Title: Plasticity of fimbrial genotype and serotype within populations of *Bordetella pertussis*: analysis by paired flow cytometry and genome sequencing

Running Title: *B. pertussis* fimbrial serotype plasticity

Contents category: Microbial Pathogenicity

Authors: Thomas E. Vaughan[1]*, Catherine B. Pratt[1], Katie Sealey[2,3], Andrew Preston[2], Norman K. Fry[3] & Andrew R. Gorringe[1]

Addresses: [1]Public Health England, Porton Down, Salisbury, SP4 0JG, United Kingdom. [2]Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom. [3] Respiratory and Vaccine Preventable Bacteria Reference Unit, Public Health England - Microbiology Reference Services, Colindale, 61 Colindale Avenue, London, NW9 5EQ, United Kingdom.

***Author for correspondence:** Thomas E. Vaughan, thomas.vaughan@phe.gov.uk, Tel. +44 1980 612421, Fax +44 1980 611310

Number of words: 189 (Abstract), 5285 (Main Text)

Number of tables 4

Number of figures 2

Abstract

The fimbriae of *Bordetella pertussis* are required for colonisation of the human respiratory tract. Two serologically distinct fimbrial subunits, Fim2 and Fim3, having been considered important vaccine components for many years, are included in the Sanofi Pasteur 5-component acellular pertussis vaccine, and the World Health Organisation recommends the inclusion of strains expressing both fimbrial serotypes in whole cell pertussis vaccines. Each of the fimbrial major subunit genes, *fim2*, *fim3*, and also *fimX*, has a promoter poly(C) tract upstream of its −10 box. Such monotonic DNA elements are susceptible to changes in length via slipped-strand mispairing *in vitro* and *in vivo*, which can potentially cause on/off switching of genes at every cell division. Here we describe intra–culture variability in poly(C) tract lengths and in the resulting fimbrial phenotypes in 22 recent United Kingdom *B. pertussis* isolates. Owing to the highly plastic nature of the fimbrial promoters, we used the same cultures for both genome sequencing and flow cytometry. Individual cultures of *B. pertussis* each contained multiple fimbrial serotypes and multiple different fimbrial promoter poly(C) tract lengths, which supports earlier serological evidence that *B. pertussis* expresses both serotypes during infection.

Introduction

Bordetella pertussis causes whooping cough and this continues to be a public health problem despite high vaccine coverage. Rates of pertussis disease have increased in many countries in recent years and this is thought to be due to a combination of strain evolution (Mooi *et al*., 2013) and faster waning immunity provided by acellular pertussis (aP) vaccines compared to whole cell pertussis vaccines (wP) used previously (Tartof *et al*., 2013). In addition, evidence from a baboon model of pertussis indicates that acellular vaccines may be less effective than whole cell vaccines at preventing acquisition and transmission of the organism (Warfel *et al*., 2014). It is important for the pathogenesis of pertussis disease that *B. pertussis* is able to adhere to the mucosal surface in the upper respiratory tract via a number of adhesins including filamentous haemagglutinin, pertactin and fimbriae (Mattoo & Cherry, 2005). *B. pertussis* expresses two serologically distinct fimbriae which are composed of either Fim2 or Fim3 subunits (22.5 and 22.0 kDa respectively). Subtypes of Fim2 (*fim2-1* and *fim2-2*) and Fim3 (*fim3-1*, *fim3-2* and *fim3-3*) have also been described (Packard *et al*., 2004; Tsang *et al*., 2004). Both Fim2 and Fim3 subunits are assembled in a coil to produce long filaments on the surface of the bacteria (Heck *et al*., 1996). Fimbriae (co-purified Fim2/3) have been shown to be protective antigens in mouse models of pertussis (Heininger *et al*., 1998; Robinson *et al*., 1989a) and are components of an acellular vaccine which showed high efficacy in phase III clinical trials (Gustafsson *et al*., 1996; Olin, 1997). Household contact studies performed during these trials showed a correlation between antibody responses to fimbriae and protection against disease (Cherry *et al*., 1998; Storsaeter *et al*., 1998).

Expression of many *B. pertussis* virulence-related genes, including *fim2* and *fim3*, is regulated by the BvgA/S two-component system. (Decker *et al*., 2012). Transcription of *fim3* is proposed to occur when a dimer of phosphorylated BvgA is bound to the *fim3* promoter's −35 region at the same time as region 4 of the RNA polymerase σ^{80} subunit and the C-terminal domain of an RNA polymerase α subunit (Decker *et al*., 2011). Poly(C) tracts, and homopolymeric tracts generally, are

under-represented in the *B. pertussis* chromosome (Coenye & Vandamme, 2005), and the possibility of rapid antigenic variation via slipped-strand mispairing at these sites has received limited attention. However, each of the *fim* genes possesses a poly(C) tract between its −35 and −10 promoter regions. Transcription-permissive lengths for the fimbrial poly(C) tracts have been determined (Willems *et al*., 1990; Chen *et al*., 2010), but the frequency of slipped-strand mispairing within them has not been previously described. The fimbrial poly(C) tracts of *B. pertussis* have a notable counterpart in the *porA* promoter of *Neisseria meningitidis*, where altered transcription levels, caused by changes in the length of a poly(G) tract between the −35 and −10 regions, can facilitate the evasion of bactericidal anti-PorA antibodies (Tauseef *et al*., 2013).

In this study, we have analysed whole genome sequencing reads from a panel of 96 *B. pertussis* strains isolated in the United Kingdom between 1920 and 2012, and shown that cultures of *B. pertussis* contain multiple subpopulations with differing fimbrial poly(C) tract lengths. In addition, we preserved aliquots from 22 of the cultures used for genome sequencing and have analysed Fim2 and Fim3 expression on individual bacteria using labelled anti-Fim2 and anti-Fim3 monoclonal antibodies in two-colour flow cytometry. This demonstrated that bacteria expressing Fim2 and/or Fim3 are present in cultures previously shown by standard slide agglutination methods to express only one fimbrial serotype. Thus, these data show that expression of Fim2 and Fim3 by *B. pertussis* is more flexible than generally acknowledged, and this flexibility may help to explain the rapid shifts in predominant serotype observed in many countries.

Materials and methods

Bacteria

Ninety-six *B. pertussis* isolates UK001 to UK06, UK008 to UK011, UK014 to UK018, UK020 to UK100 were obtained from the Respiratory and Vaccine Preventable Bacteria Reference Unit (RVPBRU), Public Health England (PHE) Colindale, London. The majority of these isolates (*n* =

87) were submitted by hospital laboratories as part of the enhanced surveillance of pertussis (Van Buynder *et al*., 1999). Additional isolates were previously obtained from the Wellcome Bacterial Collection, held by the National Collection of Type Cultures (NCTC) (n = 4); the NCTC (n = 2) and the former Centre for Applied Microbiology and Research, now PHE Porton Down, Salisbury (n = 3) (Fry *et al*., 2001).

Following culture for 48 to 72 h on blood charcoal agar (PHE Media Services, Colindale), *B. pertussis* isolates were serotyped using a slide agglutination assay with rabbit polyclonal antisera against antigens 1, 2, and 3 (89/596, 89/598, and 89/600, respectively; National Institute for Biological Standards and Controls, Potters Bar, United Kingdom). Antigen 1 (agglutinogen 1) is expressed by all isolates of *B. pertussis*, thus the notation Fim−, Fim2, Fim3 and Fim2,3 (Hallander *et al*., 2005) is used instead of the historical notation serotype 1, serotype 1,2 serotype 1,3 and serotype 1,2,3 (Robinson *et al*., 1989b) as previously described (Litt *et al*., 2009). The terms Fim− and Fim2−Fim3− are used interchangeably in the text. Cultures for this study were grown on charcoal agar for 72h at 37^{o}C. For flow cytometry analysis, bacteria were resuspended in PBS containing 1% formaldehyde and the OD_{600} nm adjusted to 2.0.

Genome sequencing

Genome sequencing of UK isolates UK001 to UK100 was performed at the Sanger Institute and detailed analysis of these sequences will be the subject of a paper in preparation. Briefly, paired-end libraries were prepared from genomic DNA, amplified with a KAPA HiFi library amplification kit (Kapa Biosystems Inc, Wilmington MA), and sequenced as 150 bp paired-end reads using a MiSeq desktop sequencer (Illumina Inc, San Diego, CA). The unassembled short reads are available in fastq.gz format from the Sanger Institute's ftp site. Locations of the files are in Table S1 (supplementary materials). Additional genome sequence files (non-UK isolates) are listed in Table S2 (supplementary materials).

Extraction of poly(C)-tract–spanning reads from fastq.gz files

Each fastq.gz file was read with *kseq_test*, a minimal example program for the `kseq.h` fastq parser library (Li, 2011; Li, 2012). Output from *kseq_test* was piped into a script written in AWK (Aho *et al*., 1988), within which the sequence field of each Illumina read was put through two successive filters.

The first filter compared each read to the set of all plausible fimbrial poly(C) tracts (6–20 base pairs) flanked by a minimal *B. pertussis* Tohama I–derived context (Parkhill *et al*., 2003). The search patterns used in this first filter are shown as regular expressions in Table 1. The *fim2* regular expressions contain a `[CT]` character list, corresponding to the deoxythymidine that punctuates the poly(C) tract of *B. pertussis* Tohama. Three regular expressions are necessary to prevent *fim3* and *fimX* sequences from being misidentified as each other. Reverse-complement patterns, not shown in Table 1, were also used, giving a total of 308 patterns upon evaluation of the regular expressions. Non-matching reads (including all reads ending within poly(C) tracts) were discarded, and exact matches were tested in the second filter.

The second filter took the segments of each read outside the initial match, joined them and aligned the resulting chimaera to the equivalent joined 30 bp segments from *B. pertussis* Tohama I. For this alignment, EMBOSS *needle* (Rice *et al*., 2000) was used via AWK's `getline`. The alignment was forced to be local with respect to query and global with respect to reference (by invoking *needle* with an endweight, a zero end-open penalty and high gap open/extend penalties). Reads were given a positive plausibility score if this second filter resulted in a plausible alignment (aligned join-sites, ≥ 66.7 % identity across ≥ 10 bp where flanking sequence was available on both sides of the initial match, ≥ 80 % identity across ≥ 20 bp where flanking sequence was available on only one side of the initial match); otherwise reads were given a negative score. Each fastq record was reassembled

into a one-line tab-delimited format and augmented with gene name (*fim2*, *fim3* or *fimX*), isolate name, fastq.gz file name and plausibility score.

The *kseq_test* preprocessing step and the AWK/*needle* double filter were combined as a single Bourne shell script, and summary counts were extracted from the tab-delimited output using a second shell script (supplementary materials: `MeasurePolyC.sh` and `FastqplusToTable.sh`). The AWK script embedded in `MeasurePolyC.sh` induced a memory leak in GNU *awk*, but not in *nawk* or *mawk*. Arithmetic means and standard deviations (Table 3 and supplementary materials, `CoverageStatsUKstrains.awk`) were calculated using the Welford–Knuth algorithm (Welford, 1962; Knuth, 1997, pp. 231–232).

Detection of Fim2 and Fim3 by flow cytometry

200 µl of each suspension (OD_{600} = 0.1) of plate-grown bacteria was added to a well of a 96-well microtitre plate. The plate was centrifuged at 3060 × g for 5 min. and supernatant was removed. Fim2- and Fim3-specific monoclonal antibodies (06/124 and 06/128 respectively, obtained from the National Institute of Biological Standards and Control, UK) were bound to AlexaFluor 488 and Alexafluor 647, respectively, using a Zenon® Alexa Fluor® Mouse IgG_1 Labelling Kit (Life Technologies, Paisley, UK). 100 µl of each antibody at a dilution of 1 in 5000 was added to each test well and pellets were resuspended. The plate was incubated at 25 °C with shaking, for 30 min. The plate was centrifuged as above and washed, and pellets were resuspended in 200 µl blocking buffer (PBS, 1 % BSA). Samples were analysed using a Beckman Coulter CyanADP flow cytometer, and Summit® software was used to calculate the median fluorescence of each fluorophore for each sample. Unstained bacterial cells were used to verify the absence of fluorophore-independent counts, which are minimised by the use of filter-sterilised buffers and sheath fluid and by the collection of 10,000 events per sample.

A gate was drawn to include the bacteria plus conjugate control (no antibody), resulting in a median fluorescence >20 being considered positive for Fim expression. To ascertain the percentage of bacterial cells expressing Fim2 or Fim3, AlexaFluor 488 fluorescence was plotted against AlexaFluor 647 fluorescence for each sample. Gates were drawn to enclose the negative control sample and used to determine the demarcation between a positive and negative result. This gate was then applied to all test samples, allowing the determination of the percentage of events, or individual bacteria, which were not fluorescent, singly positive for AlexaFluor 488, singly positive for AlexaFluor 647 or doubly positive for both Alexafluor 488 and Alexafluor 647, thereby determining the percentage of individual bacteria expressing Fim2, Fim3 or both Fim2 and Fim3 within a population.

Results

Multiple fimbrial promoter variants within cultures

From the 96 unassembled UK *B. pertussis* genome sequences, 2648 reads were matched exactly to one of the *fim*2 regular expressions in Table 1 or to one of the reverse complements thereof. After a second filter to eliminate false positives and uncertain matches, 2580 *fim2* poly(C) tract–spanning reads were kept. The twice-filtered reads contained *fim2* poly(C) tracts ranging in length from 6 to 20 bp. Multiple *fim2* poly(C) tract lengths were found within samples (Table 2), with a mean intra–isolate range of 4.2 bp (standard deviation = 2.3 bp). The deoxythimidine residue that punctuates the *fim2* poly(C) tract of *B. pertussis* Tohama was also found in UK001, UK002 and UK004 (58%, 17% and 23% of reads respectively), and was counted as an internal deoxycytidine for the purpose of poly(C) tract length enumeration. Our numbering convention thus differs from that of Chen *et al.* (2010), who examined only T-punctuated *fim2* poly(C) tracts, and whose enumeration excluded deocytidines on the 5′ side of the punctuating deoxythymidine.

Using the equivalent double filter for *fim3*, 1914 reads were matched exactly by the first filter, and 1660 *fim3* poly(C) tract–spanning reads were kept after the second filter. The twice-filtered reads contained *fim3* poly(C) tracts ranging in length from 6 to 20 bp. As with *fim2*, multiple poly(C) tract lengths were found within samples (Table 2), with a mean intra–isolate range of 5.2 bp (standard deviation = 3.0 bp).

For *fimX*, 4463 reads were matched by the first filter, and 4381 *fimX* poly(C) tract–spanning reads were kept after the second filter. The *fimX* poly(C) tracts were less variable than those of the other two fimbrial genes (mean intra–isolate range = 1.1 bp, standard deviation = 0.4). In 77 of the 96 genome-sequenced UK isolates, *fimX* poly(C) tracts did not exceed 7 bp in length. Isolate UK071 contained a *fimX* poly(C) tract of 9 bp, the longest among the UK isolates.

Read lengths for UK025 to UK100 were all 150 bp (average of 1.26 million reads per isolate, standard deviation = 0.26 million), whereas read lengths for other available *B. pertussis* Illumina sequences were 54 bp (average of 5.40 million reads per isolate for UK strains, standard deviation = 1.66 million). For *fim2* and *fimX*, poly(C) tract–spanning reads were more numerous in the genomes sequenced at 54 bp per reads, whereas for *fim3*, poly(C) tract–spanning reads were more numerous in the genomes sequenced at 150 bp per read (Table 3). Since a 54 bp read is less likely than a 150 bp read to span a given sequence, the higher tally of *fim2* and *fimX* poly(C) tracts in the 54 bp genome sequences can be attributed to the higher coverage of those genomes (71.4-fold versus 46.1-fold for the 150 bp genome sequences). For *fim3*, the higher tally in the 150 bp genome sequences can be attributed to the broader size range of poly(C) tracts ($\overline{range}_{\text{fim3}}$ = 5.2 bp versus $\overline{range}_{\text{fim2}}$ = 4.2 bp), with the longer part of this range being more likely to be encompassed by longer reads.

After normalisation to genome coverage, the yields of *fim2* and *fim3* poly(C) tract–spanning read were higher from the 150 bp genomes than from the 54 bp genomes, but the yield for *fimX* stayed higher for the 54 bp genomes (Table 3, column 8).

Presence of multiple serotypes within cultures

In this study an aliquot of culture prepared for genome sequencing from strains UK025 to UK046 was preserved by the addition of formaldehyde to 1 %. Individual bacteria in these preserved suspensions were analysed for expression of Fim2 and Fim3 using anti-Fim2 and anti-Fim3 monoclonal antibodies and two-colour flow cytometry. The flow cytometry scatter plots obtained and the poly(C) tract lengths determined for each isolate are shown in Figure 1. In each of 22 isolates examined by two-colour flow cytometry, *B. pertussis* cells of all four possible fimbrial serotypes $Fim2^{-}Fim3^{-}$, $Fim2^{-}Fim3^{+}$, $Fim2^{+}Fim3^{-}$ and $Fim2^{+}Fim3^{+}$) were present (Fig. 1, scatterplots; Fig.2). One fimbrial serotype was dominant (>55 % of the population) in 21 out of the 22 isolates (Figure 2). In isolate UK043, which had a bvg^{-} colony morphology, $Fim2^{-}Fim3^{-}$ cells were dominant (89% of the population). FimX surface expression was not investigated owing to the lack of available antibodies.

Gates were set so that the bacteria plus conjugate control was 99% enclosed inside the $Fim2^{-}Fim3^{-}$ quadrant. Although this may misidentify some events as $Fim2^{-}Fim3^{-}$ (for instance the lower portion of the Fim3+ cloud of UK042 in Figure 1), it does ensure that false positives are avoided. The diagonal streaking that can be seen in the $Fim2^{+}Fim3^{+}$ quadrant of some scatterplots (Fig. 1) is attributable to clumping of bacteria, with single bacteria towards the lower left and clumped bacteria towards the upper right. This analysis used a gate that excluded most clumped bacteria, but when a larger gate was used so as to include more clumped bacteria, the fraction of $Fim2^{+}$ or $Fim3^{+}$ bacteria remained very similar (data not plotted). Hence the results presented here are not dependent on the chosen flow cytometry analysis gate.

In all 22 of these isolates, the *fim2* and *fim3* promoter poly(C) tracts were present in the Illumina sequencing data as multiple variants. The broadest range for *fim2* occurred in UK086, with poly(C) tract lengths ranging from 7 to 17. The broadest range for *fim3* occurred in UK057, with poly(C) tract lengths ranging from 7 to 19.

Relationship between median fimbrial poly(C) tract length and serotype

Among the 21 bvg$^+$ isolates studied by paired genome sequencing and two-colour flow cytometry, Fim2 median fluorescence was high in 8 phenotypically bvg$^+$ isolates that had median *fim2* poly(C) tracts lengths of 12 to 16 (Table 2). In the other 13 bvg$^+$ isolates tested in this paired analysis, Fim2 median fluorescence was at least 13-fold lower, and median poly(C) tract lengths were either 10 or 11. *fim2* median poly(C) tract lengths longer than 16 were not seen in the 22 isolates tested by flow cytometry but were seen in four of the other 74 UK isolates.

Fim3 median fluorescence was high in 13 phenotypically bvg$^+$ isolates that had median *fim3* poly(C) tracts lengths of 13 or 14. In the other 8 bvg$^+$ isolates tested, Fim3 median fluorescence was at least 11-fold lower, and median *fim3* poly(C) tract lengths were in the range 10 to 13. *fim3* median poly(C) tract lengths longer than 14 were not seen in the 22 isolates tested by flow cytometry but were seen in four of the other 74 UK isolates.

In UK043, *fim2* and *fim3* median poly(C) tract lengths were 15 and 14 respectively, but median fluorescence was low for both Fim2 and Fim3 Over 70% of cells were Fim2$^-$Fim3$^-$ (Fig 2). This isolate had bvg$^-$ colony morphology, and this decoupling of surface expression from poly(C) tract length was therefore expected. It was, however, serotyped by agglutination as Fim3$^+$.

Among the other 74 sequenced UK isolates (which were not analysed by flow cytometry) UK015 provided the sole example of a homogeneous *fim2* promoter (38 Illumina reads spanning the *fim2* poly(C) tract, all with a poly(C) tract length of 9 nucleotides). The range of *fim2* poly(C) tract

lengths in UK010 was not determined since only one spanning read was available. There were no unambiguous examples of homogenous *fim3* promoters (only one spanning read was available for UK017).

For Fim2 and Fim3, the relationships between promoter poly(C) tract lengths and expression are therefore as follows: a bvg^+ population is $Fim2^+$ if the median poly(C) tract length is > 11 (upper bound not established, but at least 16), and a bvg^+ population is $Fim3^+$ if the median poly(C) tract length is > 13 (upper bound not established, but at least 14).

Optimal poly(C) tract lengths for *B. pertussis* fimbrial promoter activity have been proposed from observations made with luciferase reporter gene constructs (Chen *et al*., 2010). In nineteen of the 21 bvg^+ isolates described here, surface expression of Fim2 or Fim3 correlated with a majority of *fim2* or *fim3* promoters having poly(C) tracts unambiguously in the proposed active range (Figure 1, shaded in bar charts). . However, the correlation broke down for two isolates. In UK037, the majority of *fim2* poly(C) tracts were outside the proposed optimal range, but the majority of cells in this culture were $Fim2^+$. Similarly, in UK044, the majority of *fim3* poly(C) tracts were outside the proposed optimal range, while the majority of cells were $Fim3^+$.

Relationship between poly(C) tract length and serotype as determined by agglutination

Serotypes determined by agglutination were not uniformly in agreement with poly(C) tract length distributions. Thirteen of the 96 genome-sequenced isolates (including UK043) had serotypes by agglutination that conflicted with the rules proposed above connecting median poly(C) tract length to expression (Table 4); and if upper bounds for “active” median poly(C) tract lengths could be established, additional conflicting examples would probably be apparent. For 59 of the 96 isolates, serotypes by agglutination were in agreement with median poly(C) tract lengths.

Serotypes by agglutination were in agreement with the flow cytometric measurements of Fim2 and Fim3 for twenty of the 22 isolates that were analysed by flow cytometry (Table 2). The exceptions were UK030 ($Fim2^{-}Fim3^{+}$ by agglutination, but 88% $Fim2^{+}Fim3^{-}$ by flow cytometry) and the bvg^{-} isolate UK043. Four of the 96 genome-sequenced UK isolates were serotyped as Fim2,3, and one was serotyped as Fim−, but these five were not included in the flow cytometric analysis (Supplementary Table S1: UK001, UK003, UK006, UK010 and UK068).

Subpopulations with the potential to express FimX

In 87 of the 96 UK isolates, all *fimX* poly(C) tracts were 8 bp or shorter, suggesting that these strains are $FimX^{-}$. The longest *fimX* poly(C) tract found in a UK isolate was 9 bp, in UK024, which is shorter than the proposed optimum length for expression (Chen *et al*., 2010).

Since no reagents are available to detect FimX protein, the possible existence of $FimX^{+}$ subpopulations can be tested only by scanning genome sequences for *fimX* poly(C) tracts of plausibly transcription-permissive lengths. A further 275 unassembled *B. pertussis* genome sequences, all from non–United Kingdom isolates, were therefore filtered through the same script as was used for the 96 UK isolates. In each of 195 non-UK isolates, the longest observed *fimX* poly(C) tract length was 7 bp, and the longest *fimX* poly(C) tract lengths observed in the whole dataset were 10 bp (found in two isolates), which is shorter than the proposed optimum length for expression.

Discussion

Fimbriae are required by *B. pertussis* for colonisation of the human respiratory tract; and we have shown in this study using two-colour flow cytometry, that individual cultures each contain cells of more than one fimbrial serotype. In the same cultures, we have detected multiple variants of the promoters for each of the serologically distinct fimbrial major subunits. This study therefore provides cellular and genomic corroboration for the suggestion by Heikkenen *et al.* (2008), and

recently reported serological evidence, that *B. pertussis* expresses Fim2 *and* Fim3 during infection irrespective of the predominant serotype of case isolates (Hallander *et al*., 2013).

The intra–culture serotype heterogeneity described here reflects the intrinsic susceptibility of repetitive DNA elements (in this case, promoter poly(C) tracts) to slipped-strand mispairing during DNA replication. The steric feasibility of bulge defects in the base-pairing of nucleic acids was demonstrated more than 50 years ago, and the formation of such defects by slipped-strand mispairing was postulated immediately as a mechanism for point mutations in DNA regions of low complexity (Fresco & Alberts, 1960). Insertions and deletions within runs of short (1 to 10 bp) repeated elements have subsequently been observed in human genetic disorders and in antigenic switching by pathogens (reviewed by van der Woude & Bäumler, 2004). Sporadic Huntingdon's disease arises from expansion of a run of sequential CAG glutamine codons in the *HTT* gene (Myers *et al*., 1993), and similar CAG expansions are seen in several other neurodegenerative disorders. Analogous in-frame trinucleotide repeats (AGT or TCT) occur in the MgPa adhesin operon of *Mycoplasma genitalium*, and variations in the number of repeat units have been observed within and between isolates (Ma *et al*., 2012). The opacity genes of *Neisseria gonorrhoeae* are subject to reversible frameshifts, at a rate of 10^{-4}–10^{-3}/cell/generation, resulting from changes in the number of copies of a CTCTT repeat in the signal peptide–encoding sequence (Murphy *et al*., 1989). Variation in the number of copies of a TAAA repeat in the *nadA* promoter of *Neisseria meningitidis* MC58 leads to colonies with differential NadA expression appearing on the same culture plate (Martin *et al*., 2003).

The simplest possible repeat elements, homopolymeric tracts, account for many reported instances of slipped-strand mispairing in bacteria. There are 5684 homopolymeric tracts longer than 4 bp in the *B. pertussis* Tohama I genome sequence (Parkhill *et al*., 2003), of which 78% are 5 bp in length. The 6 bp homopolymeric tracts account for 15% of the total, and frequency decreases with

increasing length. In a survey of expected versus observed homopolymeric tract frequencies in 139 sequenced bacterial genomes, *B. pertussis* was ranked twelfth lowest with an expected/observed ratio of 0.23 (Coenye & Vandamme, 2005), and most reports of homopolymeric tract–associated antigenic variation are from bacteria with higher expected/observed ratios (for instance, 2.1 in *Neisseria meningitidis* and 3.5 in *Helicobacter pylori*). It is therefore striking that the earliest report of *B. pertussis* slipped-strand mispairing was in the *bvg*S gene, where the impact is magnified by pleiotropy. Stibitz *et al.* (1989) found that bvg^- strains arose by frameshifting expansion of a C_6 tract at *bvg*S codons 1103–4, and that the frequency of the $bvg^{+/-}$ switch varied between two strains.

Transcription from the fimbrial promoters requires that their poly(C) tracts be longer (Chen *et al.*, 2010) than the footprints of some DNA polymerases (Beese *et al.*, 1993; Eom *et al.*, 1996), so the heterogeneity of fimbrial promoters reported here is perhaps unsurprising. Fazekas *et al.* (2010) were able to improve the fidelity of PCR amplification across homopolymeric tracts by using PCR enzymes with extended DNA-binding footprints. Compared to a 68°C PCR extension, DNA replication in *B. pertussis* should be less vulnerable to slipped-strand mispairing because the 3′ end of the nascent strand is less likely to dissociate at physiological temperatures. Nevertheless, the imperfect fidelity of C_6 replication in *bvg*S, along with the apparent selection against homopolymeric tracts throughout the genome, suggests that *B. pertussis* is unable to replicate homopolymeric tracts with high fidelity if they are longer than 5 bp. The fimbrial poly(C) tracts of *B. pertussis*, being several nucleotides longer, are directly analogous to the poly(C) tract of *opc* (Sarkari *et al.*, 1994) and the poly(G) tract of *porA* (van der Ende *et al.*, 1995; Tauseef *et al.*, 2013), both in *N. meningitidis*. In each of these examples, expression of a surface antigen is modulated by frequent and apparently stochastic changes in the spacing between the −35 and −10 promoter regions.

In an organism that uses slipped-strand mispairing to effect antigenic variation, cell division must, at least some of the time, result in daughter cells inheriting surface antigens that they are unable to express anew. The correlation between fimbrial poly(C) tract length distribution and surface expression is therefore likely to be imperfect, especially if antigenic variation is rapid. This may be exemplified in the present study by the two isolates with poly(C) tracts lengths predominantly outside the transcription-permissive range estimated by Chen *et al.* (2010); though an alternative explanation might be that the estimates of transcription-permissive lengths were compromised by slipped-strand mispairing in the reporter gene constructs used. In this study, we obtained statistical estimates of transcription-permissive fimbrial poly(C) tract lengths, as distinct from the previous reporter gene–based estimates, but isolates with conflicting serotype (by agglutination assay) were still apparent (Table 4). However, the objective of the present study was not to determine precise limits for promoter activity, and the range of poly(C) tract lengths in individual cultures was broad enough to encompass the multiple serotypes seen in the same cultures (Figure 1).

The major impediment to estimating *in vivo* poly(C) tract diversity lies in the potential for laboratory DNA polymerases to add to the diversity *in vitro*. In our hands, *B. pertussis* fimbrial poly(C) tracts can be sequenced successfully using dideoxy chemistry if they are amplified in *Escherichia coli* as plasmid inserts, implying correct amplification by the *E. coli* DNA polymerase. However, when PCR-amplified fimbrial promoters are sequenced without cloning, multiple superimposed traces appear on the downstream side of the poly(C) tract, representing multiple poly(C) tract lengths of which some might represent slipped-stranded mispairing in the PCR tube rather than in the organism. For the present study, we sought to avoid PCR-generated diversity by using high coverage sequence data. The 10-cycle amplification in the Kapa HiFi library amplification protocol amounts to 10000-fold less amplification that takes place in a 30-cycle PCR reaction, so the number of *in vitro*-generated poly(C) tract lengths is likely to be minimal. In

addition, only a few of the poly(C) tract length distributions in Figure 1 resemble the bell-shaped distribution that would be expected for purely PCR-generated diversity.

The distribution of *B. pertussis* fimbrial serotypes, as determined by agglutination, has varied between countries and over time, and may have been influenced by differing vaccination policies (reviewed by Poolman & Hallander, 2007). Our data indicate that the result returned by an agglutination assay is the serotype of the largest subpopulation in the sample being assayed; so the fact that temporal and geographical trends can be discerned at all suggests that *B. pertussis* populations are stably diverse, with selective pressure being required to displace a dominant serotype. Detection limits for current serotyping assays are not known, but could be determined with blended samples following separate flow cytometric analyses. The proportion of United Kingdom isolates with serotype $Fim2^+$ (by agglutination) dropped sharply from 47 % (n = 110) in 1988 to 2001, to 3 % (n = 184) in 2002 to 2004, and 2 % (n = 166) in 2005 to 2008, after the introduction of aP vaccines (a preschool booster in 2001 and a replacement for the previous wP given at 2, 3 and 4 months in 2004 (Litt *et al*., 2009), but increased in 2011–2012 (Alexander *et al*., 2012) to levels comparable with the pre-vaccine and whole-cell vaccine eras. It is possible that population immunity to the predominant serotype builds up until a tipping point is reached and the expression of the alternative serotype is favoured. However, there is no clear pattern in the median poly(C) tract lengths over these periods, partly because of a sampling bias in favour of 2012 isolates (Figure S1 supplementary materials).

FimX is 73% similar to Fim2 and 70% similar to Fim3, but has been detected only by mass spectrometry and only in one *B. pertussis* isolate (Tefon *et al*., 2011). No laboratory antibodies are available for more rapid detection, and the capacity of FimX to substitute for Fim2 and Fim3 is unknown. Most of the *fimX* poly(C) tracts counted in the present study were 6 to 8 bp long, which is probably too short to allow transcriptional initiation. The emergence of a $FimX^+$ strain from one

with exclusively short *fimX* poly(C) tracts might conceivably start with a low-probability DNA replication error followed by increasingly probable recurrences as the poly(C) tract grew further beyond the size of the DNA polymerase footprint in successive generations. The analogous phenomenon in human CAG expansion disorders, where successive generations are at increased risk as their CAG repeats expand, is termed genetic anticipation (Walker, 2007). The 10 bp *fimX* poly(C) tracts found in two strains are comparable in length to some of the *fim2* and *fim3* poly(C) tracts (Figure 1); but a robust investigation of FimX genetic anticipation requires more data than were available in the genome sequences (Table 3 column 4), and would be better accomplished by generating specific antibodies and using them in flow cytometry.

Thus, in this study we have characterised the distribution of poly(C) tract lengths in the promoters of *fim2* and *fim3* and shown that cultures contain multiple sub-populations with differing poly(C) lengths. For a subset of strains, we determined expression of Fim2 and Fim3 on each bacterium using two-colour flow cytometry, and showed that whilst there is a predominant serotype in each culture, multiple combinations of Fim2 and Fim3 expression occur in a single culture. The optimal poly(C) tract lengths for expression of Fim2 and Fim3 were greater than 11 bp or 13 bp respectively for most strains, but for 13 out of 96 strains the observed serotype conflicted with that predicted by median poly(C) tract length (Table 4). This demonstrates that expression of Fim2 and Fim3 is more fluid than previously though and confirms the serological evidence that *B. pertussis* expresses both Fim2 and Fim3 *in vivo*. It is therefore important for *B. pertussis* vaccines that rely on protection mediated by fimbriae to elicit antibodies against both fimbrial antigens.

Acknowledgements

Work at Public Health England, Porton Down has been funded by the National Institute of Health Research Centre for Health Protection Research. Katie Sealey is funded by a PhD Studentship from Public Health England.

References

Aho, A.V., Kernighan, B.W. & Weinberger, P.J. (1988). *The AWK Programming Language*, 1st edn. Addison-Wesley, Reading, MA.

Alexander, F., Matheson, M., Fry, N.K., Labram, B. & Gorringe, A.R. (2012). Antibody Responses to Individual *Bordetella pertussis* Fimbrial Antigen Fim2 or Fim3 following Immunization with the Five-Component Acellular Pertussis Vaccine or to Pertussis Disease. *Clin Vaccine Immunol* **19**, 1776–1783.

Beese, L.S., Derbyshire, V. & Steitz, T.A. (1993). Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260**, 352–355.

Chen, Q., Decker, K.B., Boucher, P.E., Hinton, D. & Stibitz, S. (2010). Novel architectural features of *Bordetella pertussis* fimbrial subunit promoters and their activation by the global virulence regulator BvgA. *Mol Microbiol* **77**, 1326–1340.

Cherry, J.D., Gornbein, J., Heininger, U. & Stehr, K. (1998). A search for serologic correlates of immunity to Bordetella pertussis cough illnesses. *Vaccine* **16**, 1901–1906.

Coenye, T. & Vandamme, P. (2005). Characterization of mononucleotide repeats in sequenced prokaryotic genomes. *DNA Res* **12**, 221–233.

Decker, K.B., Chen, Q., Hsieh, M.-L., Boucher, P., Stibitz, S. & Hinton, D.M. (2011). Different requirements for σ Region 4 in BvgA activation of the *Bordetella pertussis* promoters P_{fim3} and P_{fhaB}. *J Mol Biol* **409**, 692–709.

Decker, K.B., James, T.D., Scott, S. & Hinton, D.M. (2012). The *Bordetella pertussis* model of exquisite gene control by the global transcription factor BvgA. *Microbiology* **158**, 1665–1676.

van der Ende, A., Hopman, C.T., S., Z., Essink, B.B., Berkhout, B. & Dankert, J. (1995). Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the −10 and −35 regions of the promoter. *J Bacteriol* **177**, 2475–2480.

Eom, S.H., Wang, J. & Steitz, T.A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. *Nature* **382**, 278–281.

Fazekas, A., Steeves, R. & Newmaster, S. (2010). Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques* **48**, 277–285.

Fresco, J.R. & Alberts, B.M. (1960). The accommodation of noncomplementary bases in helical polyribouncleotides and deoxyribonucleic acids. *Proc Natl Acad Sci U S A* **46**, 311–321.

Fry, N.K., Neal, S., Harrison, T.G., Miller, E., Matthews, R. & George, R.C. (2001). Genotypic variation in the *Bordetella pertussis* virulence factors pertactin and pertussis toxin in historical and recent clinical isolates in the United Kingdom. *Infect Immun* **69**, 5520–5528.

Gustafsson, L., Hallander, H.O., Olin, P., Reizenstein, E. & Storsaeter, J. (1996). A controlled trial of a two-component acellular, a five-component acellular, and a whole-cell pertussis vaccine. *N Engl J Med* **334**, 349–355.

Hallander, H., Advani, A., Alexander, F., Gustafsson, L., Ljungman, M., Pratt, C., Hall, I. & Gorringe, A.R. (2013). Antibody responses in Sweden to individual *Bordetella pertussis* fimbriae serotype 2 or 3 following immunization with whole cell or two- or five-component acellular pertussis vaccines and following pertussis disease and in children in 1997 and 2007. *Clin Vaccine Immunol* **21**, 165–173.

Hallander, H.O., Advani, A., Donnelly, D., Gustafsson, L. & Carlsson, R.-M. (2005). Shifts of *Bordetella pertussis* variants in Sweden from 1970 to 2003, during three periods marked by different vaccination programs. *J Clin Microbiol* **43**, 2856–2865.

Heck, D.V., Trus, B.L. & Steven, A.C. (1996). Three-dimensional structure of *Bordetella pertussis* fimbriae. *J Struct Biol* **116**, 264–269.

Heikkinen, E., Xing, D.K., Olander, R.-M., Hytönen, J., Viljanen, M.K., Mertsola, J. & He, Q. (2008). *Bordetella pertussis* isolates in Finland: serotype and fimbrial expression. *BMC Microbiol* **8**, 162.

Heininger, U., Stehr, K. & Cherry, J.D. (1998). The efficacy of a whole cell pertussis vaccine and fimbriae against *Bordetella pertussis* and *Bordetella parapertussis* infections in respiratory mouse model. *Vaccine* **16**, 1255.

Knuth, D.E. (1997). Seminumerical Algorithms. Addison-Wesley Longman, Boston, MA, USA , p. 232.

Li, H. (2011). kseq_test.c. https://github.com/attractivechaos/klib/blob/master/test/kseq_test.c

Li, H. (2012). kseq.h. https://github.com/attractivechaos/klib/blob/master/test/kseq.h

Litt, D.J., Neal, S.E. & Fry, N.K. (2009). Changes in genetic diversity of the *Bordetella pertussis* population in the United Kingdom between 1920 and 2006 reflect vaccination coverage and emergence of a single dominant clonal type. *J Clin Microbiol* **47**, 680–688.

Ma, L., Jensen, J.S., Mancuso, M., Hamasuna, R., Jia, Q., McGowin, C.L. & Martin, D.H. (2012). Variability of trinucleotide tandem repeats in the MgPa operon and its repetitive chromosomal elements in *Mycoplasma genitalium*. *J Med Microbiol* **61**, 191–197.

Martin, P., van de Ven, T., Mouchel, N., Jeffries, A.C., Hood, D.W. & Moxon, E.R. (2003). Experimentally revised repertoire of putative contingency loci in *Neisseria meningitidis* strain MC58: evidence for a novel mechanism of phase variation. *Mol Microbiol* **50**, 245–257.

Mattoo, S. & Cherry, J.D. (2005). Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to *Bordetella pertussis* and other *Bordetella* subspecies. *Clin Microbiol Rev* **18**, 326–382.

Mooi, F.R., Van Der Maas, N.A. & De Melker, H.E. (2013). Pertussis resurgence: waning immunity and pathogen adaptation — two sides of the same coin. *Epidemiol Infect*, 1–10.

Murphy, G.L., Connell, T.D., Barritt, D.S., Koomey, M. & Cannon, J.G. (1989). Phase variation of gonococcal protein II: regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence. *Cell* **56**, 539–547.

Myers, R.H., MacDonald, M.E., Koroshetz, W.J., Duyao, M.P., Ambrose, C.M., Taylor, S.A., Barnes, G., Srinidhi, J., Lin, C.S. & Whaley, W.L. (1993). *De novo* expansion of a $(CAG)_n$ repeat in sporadic Huntington's disease. *Nat Genet* **5**, 168–173.

Olin, P. (1997). Efficacy trial of acellular pertussis vaccines; trial I. *Dev Biol Stand* **89**, 52–54.

Packard, E.R., Parton, R., Coote, J.G. & Fry, N.K. (2004). Sequence variation and conservation in virulence-related genes of *Bordetella pertussis* isolates from the UK. *J Med Microbiol* **53**, 355–365.

Parkhill, J., Sebaihia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T., Churcher, C.M., Bentley, S.D. & other authors (2003). Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* **35**, 32–40.

Poolman, J.T. & Hallander, H.O. (2007). Acellular pertussis vaccines and the role of pertactin and fimbriae. *Expert Rev Vaccines* **6**, 47–56.

Rice, P., Longden, I. & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276–7.

Robinson, A., Gorringe, A.R., Funnell, S.G. & Fernandez, M. (1989a). Serospecific protection of mice against intranasal infection with *Bordetella pertussis*. *Vaccine* **7**, 321–324.

Robinson, A., Ashworth, L.A. & Irons, L.I. (1989b). Serotyping *Bordetella pertussis* strains. *Vaccine* **7**, 491–494.

Sarkari, J., Pandit, N., Moxon, E.R. & Achtman, M. (1994). Variable expression of the Opc outer membrane protein in *Neisseria meningitidis* is caused by size variation of a promoter containing poly-cytidine. *Mol Microbiol* **13**, 207–217.

Stibitz, S., Aaronson, W., Monack, D. & Falkow, S. (1989). Phase variation in *Bordetella pertussis* by frameshift mutation in a gene for a novel two-component system. *Nature* **338**, 266–269.

Storsaeter, J., Hallander, H.O., Gustafsson, L. & Olin, P. (1998). Levels of anti-pertussis antibodies related to protection after household exposure to *Bordetella pertussis*. *Vaccine* **16**, 1907–1916.

Tartof, S.Y., Lewis, M., Kenyon, C., White, K., Osborn, A., Liko, J., Zell, E., Martin, S., Messonnier, N.E., Clark, T.A. & Skoff, T.H. (2013). Waning immunity to pertussis following 5 doses of DTaP. *Pediatrics* **131**, e1047–e1052.

Tauseef, I., Ali, Y.M. & Bayliss, C.D. (2013). Phase variation of PorA, a major outer membrane protein, mediates escape of bactericidal antibodies by *Neisseria meningitidis*. *Infect Immun* **81**, 1374–1380.

Tefon, B.E., Maaß, S., Özcengiz, E., Becher, D., Hecker, M. & Özcengiz, G. (2011). A comprehensive analysis of *Bordetella pertussis* surface proteome and identification of new immunogenic proteins. *Vaccine* **29**, 3583–3595.

Thompson, K. & Ritchie, D.M. (1971). *UNIX Programmers Manual*, 1st edn. Bell Laboratories, New York.

Tsang, R.S., Lau, A.K., Still, M.L., Halperin, S.A., Van Caeseele, P., Jamieson, F. & Martin, I.E. (2004). Polymorphisms of the fimbria *fim*3 gene of *Bordetella pertussis* strains isolated in Canada. *J Clin Microbiol* **42**, 5364–5367.

Van Buynder, P.G., Owen, D., Vurdien, J.E., Andrews, N.J., Matthews, R.C. & Miller, E. (1999). *Bordetella pertussis* surveillance in England and Wales: 1995–7. *Epidemiol Infect* **123**, 403–411.

Walker, F.O. (2007). Huntington's disease. *Lancet* **369**, 218–228.

Warfel, J.M., Zimmerman, L.I. & Merkel, T.J. (2014). Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model. *Proc Natl Acad Sci U S A* **111**, 787–792.

Welford, B.P. (1962). Note on a Method for Calculating Corrected Sums of Squares and Products. *Technometrics* **4**, 419–420.

Willems, R., Paul, A., Heide, H.G., Avest, A.R. & Mooi, F.R. (1990). Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *EMBO J* **9**, 2803–2809.

van der Woude, M.W. & Bäumler, A.J. (2004). Phase and antigenic variation in bacteria. *Clin Microbiol Rev* **17**, 581–611.

Tables

Table 1

Gene		Regular expression
fim2	(1)	`gccc[ct]c\{2,17\}taagacc`
	(2)	`gaatcagccc[ct]c\{2,17\}t`
fim3	(1)	`gc\{6,20\}ggacctg`
	(2)	`ccatcagc\{6,20\}gg`
	(3)	`accatcagc\{6,20\}g`
fimX	(1)	`gc\{6,20\}gaggcgt`
	(2)	`ccatcagc\{6,20\}ga`
	(3)	`tccatcagc\{6,20\}g`

Table 1. Patterns used in the first of two sequential text filters to identify genome sequence reads spanning the promoter poly(C) tracts of *fim2*, *fim3* or *fimX*. The patterns are printed in ED regular expression syntax (Thompson & Ritchie, 1971). The reverse-complement of each pattern was also used.

Table 2

Isolate	Year of isolation	Serotype by agglutination	ENA identifier	Fim2				Fim3			
				No. of reads	Median poly(C) tract length	Range of poly(C) tract lengths	Median fluorescence	No. of reads	Median poly(C) tract length	Range of poly(C) tract lengths	Median fluorescence
UK025	2008	Fim3	ERS176852	33	10	3 (9 to 11)	4.6	8	14	6 (13 to 18)	320.2
UK026	2008	Fim3	ERS176853	28	10	2 (10 to 11)	4.3	12	14	3 (13 to 15)	251.4
UK027	2008	Fim3	ERS176854	24	10	3 (9 to 11)	4.6	15	14	4 (12 to 15)	335.3
UK028	2009	Fim3	ERS176855	27	10	3 (9 to 11)	4.6	10	14	4 (12 to 15)	360.7
UK029	2009	Fim2	ERS176856	6	16	4 (14 to 17)	708.8	20	13	5 (11 to 15)	12.8
UK030	2011	Fim3	ERS176857	4	16	3 (15 to 17)	703.3	26	10	3 (9 to 11)	8.1
UK031	2011	Fim2	ERS176858	9	15	4 (13 to 16)	666.5	24	10	3 (9 to 11)	8.1
UK032	2011	Fim3	ERS176859	39	11	3 (10 to 12)	4.6	22	14	8 (12 to 19)	209.8
UK033	2011	Fim3	ERS176860	30	10	2 (9 to 10)	4.6	5	14	2 (14 to 15)	246.7
UK034	2011	Fim2	ERS176861	6	15	4 (13 to 16)	608.7	20	11	3 (10 to 12)	9.3
UK035	2012	Fim2	ERS176862	11	15	7 (11 to 17)	361	21	12	3 (11 to 13)	12
UK036	2012	Fim3	ERS176863	31	10	3 (9 to 11)	4.1	6	14	2 (13 to 14)	666.5
UK037	2012	Fim2	ERS176864	12	14	6 (11 to 16)	255.8	38	10	3 (9 to 11)	9.9
UK038	2012	Fim3	ERS176865	30	10	3 (9 to 11)	4.9	17	14	8 (12 to 19)	394.6
UK039	2012	Fim3	ERS176866	28	10	2 (9 to 10)	4.6	11	14	9 (12 to 20)	341.5
UK040	2012	Fim2	ERS176867	11	15	7 (13 to 19)	270	37	12	8 (11 to 18)	9.9
UK041	2012	Fim2	ERS176868	7	15	5 (13 to 17)	270.2	27	12	5 (11 to 15)	9.6
UK042	2012	Fim3	ERS176869	28	10	2 (10 to 11)	4.7	6	14	3 (12 to 14)	178.3
UK043	2012	Fim3	ERS176870	11	15	8 (9 to 16)	4.6	11	14	3 (13 to 15)	15.3
UK044	2012	Fim3	ERS176871	36	10	3 (9 to 11)	4.5	6	13.5	8 (11 to 18)	291
UK045	2012	Fim3	ERS176872	35	10	3 (10 to 12)	4.2	5	14	3 (13 to 15)	312.8
UK046	2012	Fim3	ERS176873	43	10	4 (9 to 12)	4	9	14	6 (12 to 17)	191.9

Table 2. Promoter poly(C) tract lengths and surface expression of Fim2 and Fim3 in *B. pertussis* isolates UK025 to UK046. ENA identifiers are European Nucleotide Archive sample names. Serotype by agglutination uses the nomenclature of Hallander *et al.* (2005). Each isolate was cultured twice: once for serotyping by agglutination and once for the other two analyses, with cultures being split between flow cytometry and whole genome sequencing. Values (apart from fluorescence) for the other 74 UK isolates are in Table S3 (supplementary materials).

Table 3

Isolates	Read length	No. of reads	Genome coverage	No. of reads spanning poly(C) tracts		Yield of reads spanning poly(C) tract		Gene
				1st filter	2nd filter	1st filter	2nd filter	
UK001 to UK024 (20 isolates)	54	5401318 *(1704184)*	71.4-fold *(22.5-fold)*	33.95 *(33.30)*	32.70 *(32.56)*	45.91% *(40.75%)*	44.22% *(39.73%)*	*fim2*
				18.50 *(15.40)*	15.10 *(14.14)*	24.93% *(18.72%)*	15.10 *(14.14)*	*fim3*
				80.15 *(33.26)*	79.40 *(32.98)*	109.70% *(21.94%)*	108.66% *(22.04%)*	*fimX*
UK025 to UK100 (76 isolates)	150	1255255 *(259298)*	46.1-fold *(9.5-fold)*	25.91 *(16.18)*	25.34 *(16.12)*	56.71% *(33.85%)*	55.45% *(33.80%)*	*fim2*
				20.32 *(14.25)*	17.87 *(13.83)*	43.56% *(26.83%)*	38.25% *(25.89%)*	*fim3*
				37.63 *(11.54)*	36.75 *(11.26)*	81.30% *(16.52%)*	79.46% *(16.47%)*	*fimX*

Table 3. Availability of Illumina reads spanning the fimbrial poly(C) tracts in *B. pertussis* isolates UK001 to UK024 and UK025 to UK100. All values other than read length are arithmetic means per isolate (standard deviations *italicised*). Yields are expressed as percentages of genome coverage. Genome sequences were not available for UK007, UK012, UK013 and UK019.

Table 4

Isolate	Serotype		Median poly(C) tract length	
	Observed	Predicted	*fim2*	*fim3*
UK001	Fim–	Fim**2**	14	9
UK003	Fim2,**3**	Fim2	16	6
UK005	Fim**3**	Fim**2**	17	13
UK010	Fim2,**3**	Fim2	17	10.5
UK011	Fim**3**	Fim**2**	14	13
UK015	Fim**3**	Fim–	9	10.5
UK030	Fim**3**	Fim**2**	16	10
UK043	Fim3	Fim**2**,3	15	14
UK051	Fim**3**	Fim**2**	12	13
UK061	Fim**3**	Fim–	11	8
UK064	Fim3	Fim**2**,3	14	14
UK072	Fim**3**	Fim**2**	12	13
UK081	Fim**3**	Fim**2**	16	9

Table 4. *B. pertussis* isolates with serotype by agglutination conflicting with serotype predicted from fim2 and fim3 median poly(C) tract length. Observed, by agglutination; Predicted, from median poly(C) tract lengths. Serotype nomenclature is as described by Hallander *et al.* (2005). Discrepant fimbrial serotypes between those observed and predicted are emboldened.

Figure legends

Figure 1. Bar charts: distribution of *fim2* and *fim3* promoter poly(C) tract lengths in unassembled genome sequencing reads for 22 United Kingdom isolates of *B. pertussis*. Scatterplots: surface expression of Fim2 and Fim3 in aliquots of the same cultures as were used for the genome sequencing, determined by two-colour flow cytometry. The dominant serotype by agglutination in each sample is stated in parentheses in the bar chart titles. Shaded bars indicate poly(C) tract lengths that were previously reported to be transcription-permissive (Chen *et al*., 2010). *fim2* poly(C) tract lengths include the C residues that precede the T residue that interrupts sequences analysed by Chen *et al.* The particles on the outside bounds of the scatterplots are those with a very high fluorescence that exceeds the dynamic range of the flow cytometer: they are more likely to be clumps of AlexaFluor-conjugated antibody than to be bacteria.

Figure 2. Relative abundances of four fimbrial serotype permutations in individual cultures of 22 UK isolates of *B. pertussis* analysed by two-colour flow cytometry.

Figure 1 (part 1 of 5)

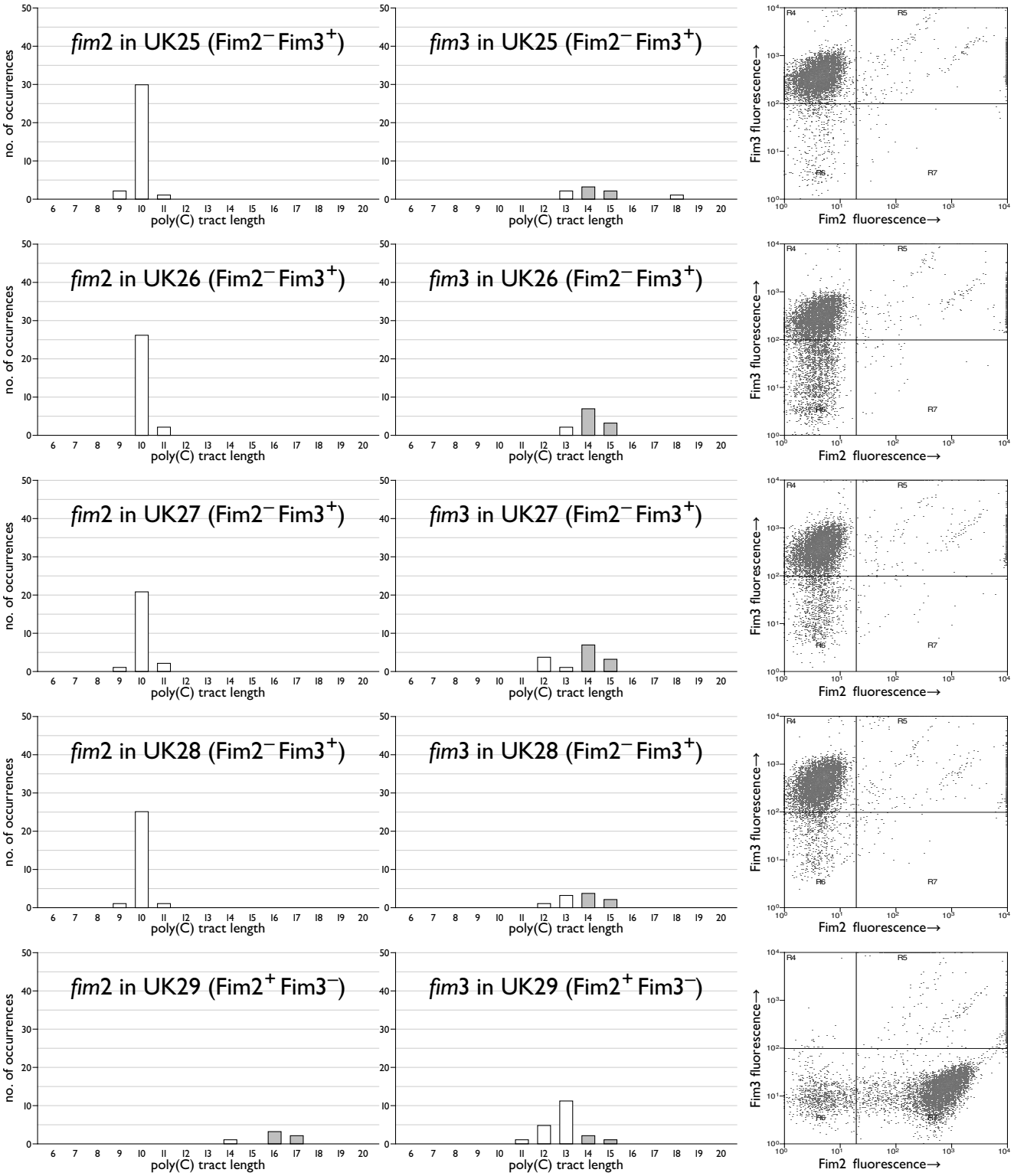


Figure 1 (part 2 of 5)

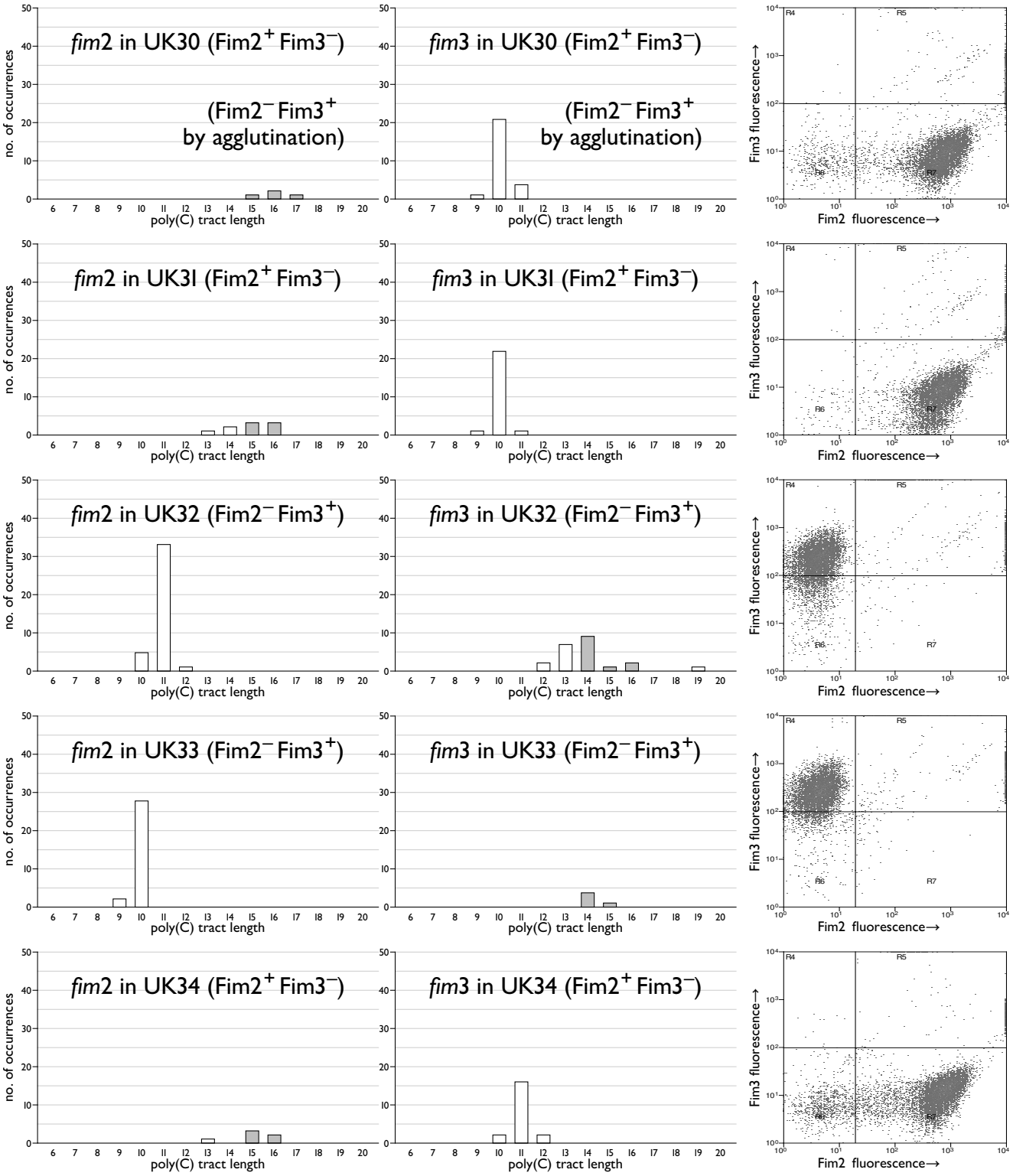


Figure 1 (part 3 of 5)

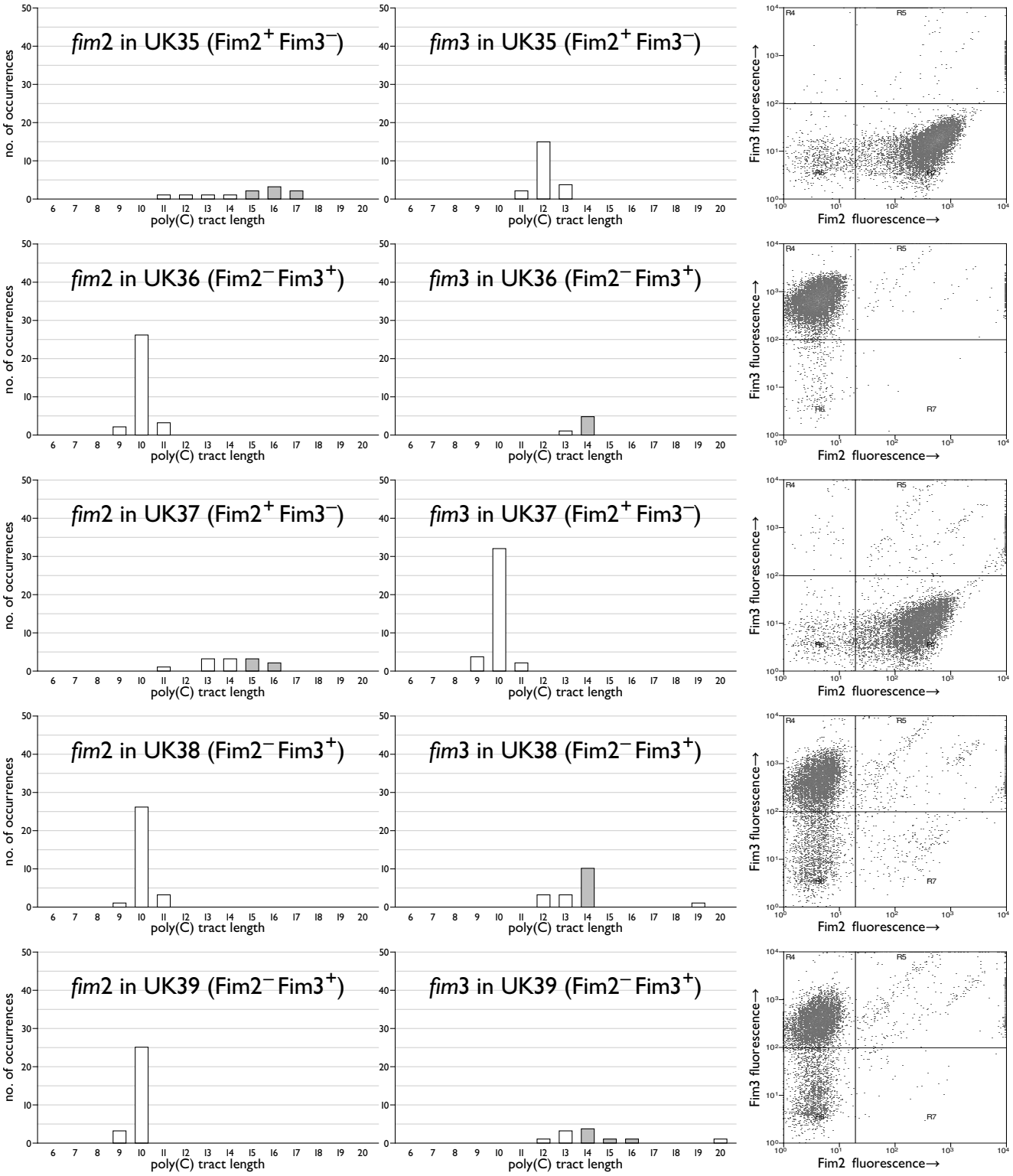


Figure 1 (part 4 of 5)

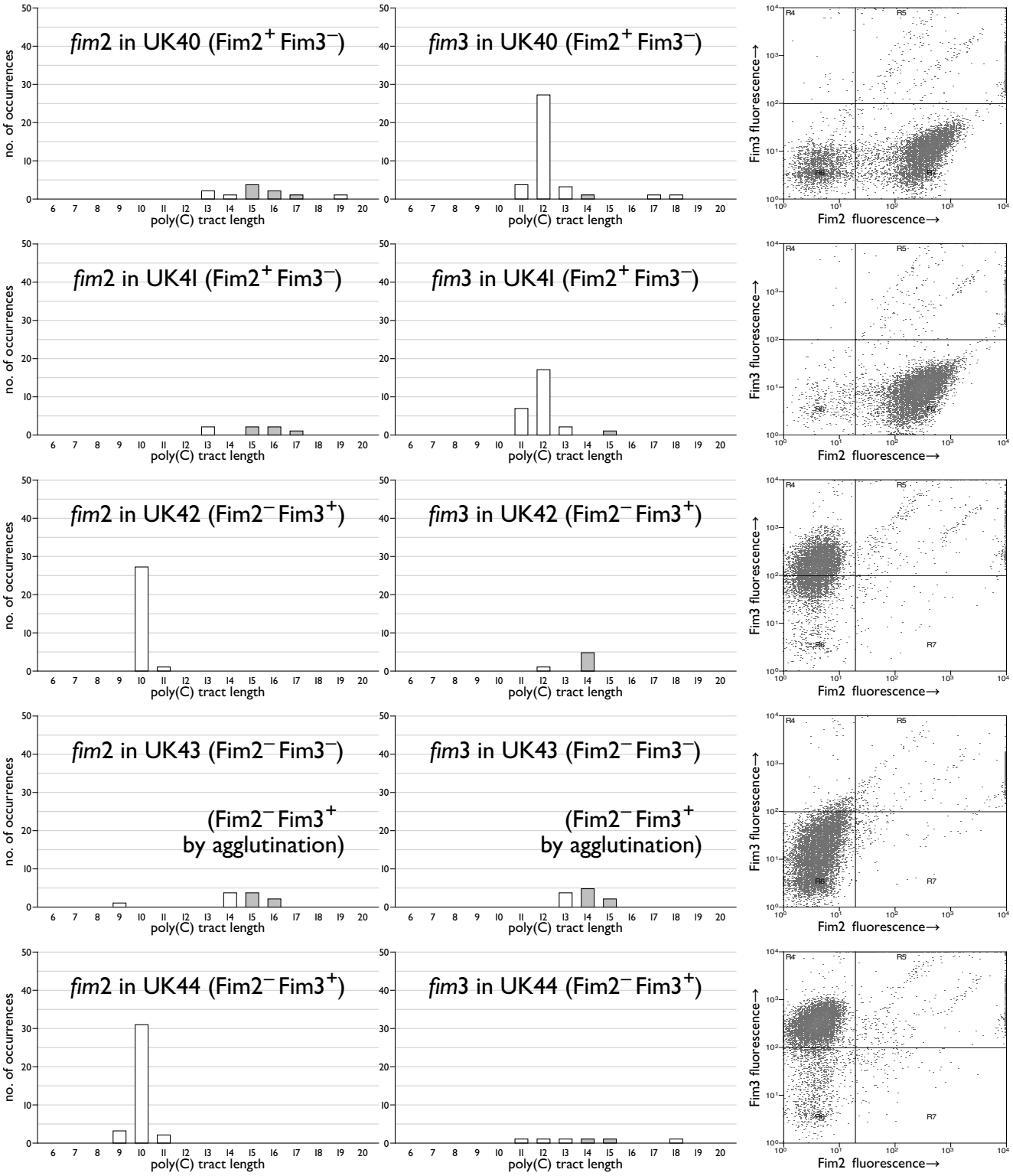


Figure 1 (part 5 of 5)

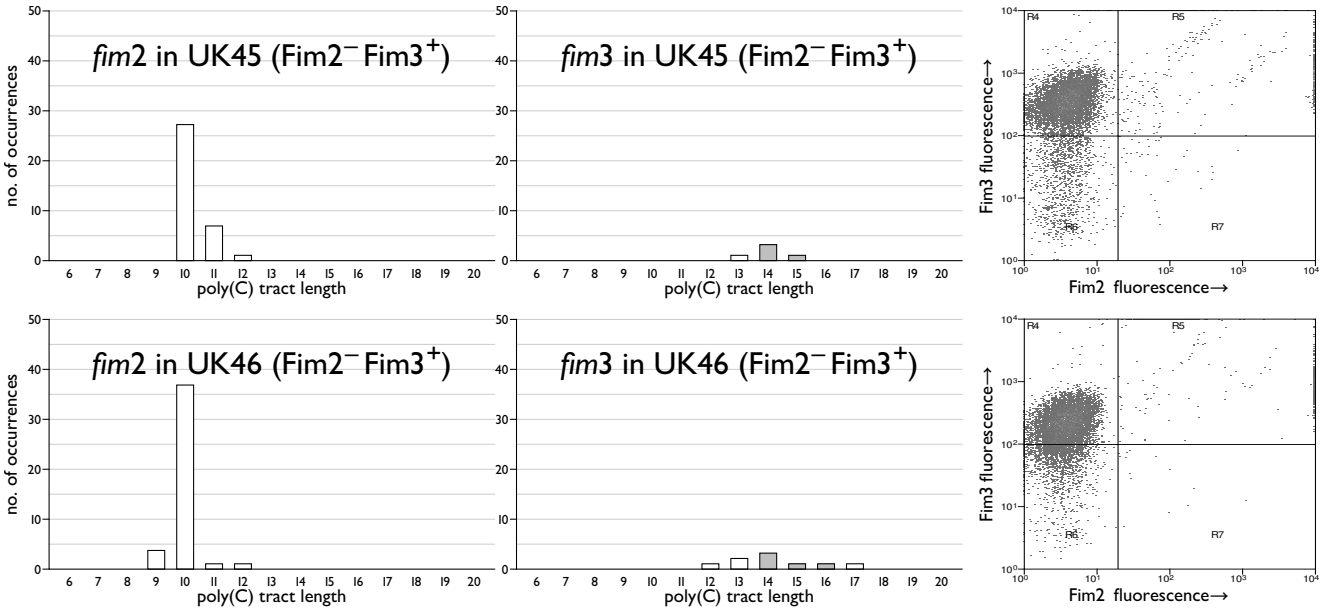
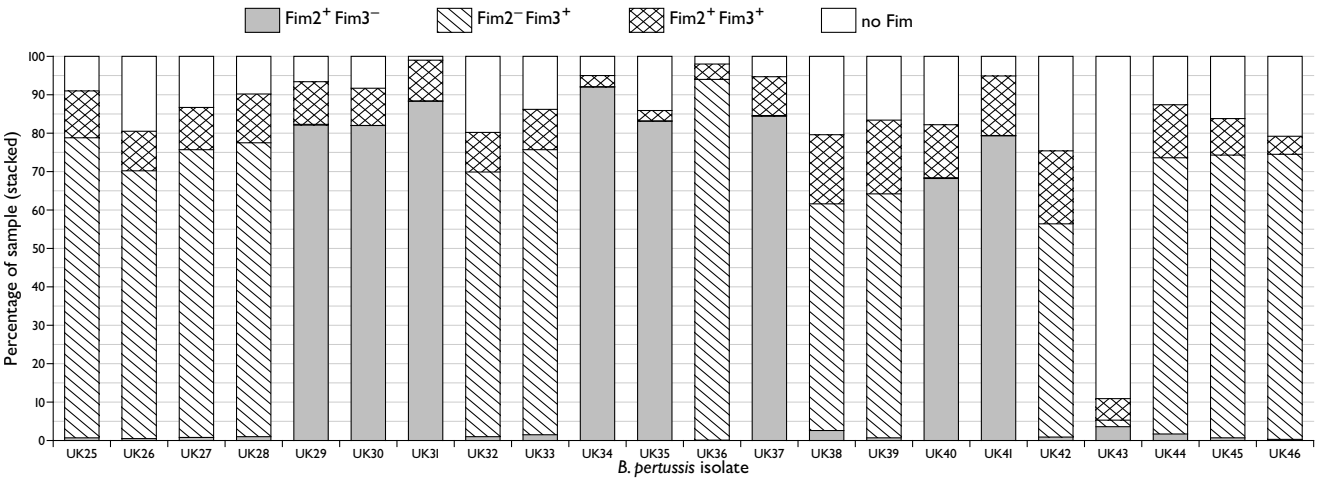


Figure 2



Supplementary tables and figures for

Plasticity of fimbrial genotype and serotype within populations of *Bordetella pertussis*: analysis by paired flow cytometry and genome sequencing

Thomas E. Vaughan[1]*, Catherine B. Pratt[1], Katie Sealey[2,3],
Andrew Preston[2], Norman Fry[3] & Andrew R. Gorringe[1]

[1]Public Health England, Porton Down, Salisbury, SP4 0JG, United Kingdom. [2]Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom. [3]Respiratory and Vaccine Preventable Bacteria Reference Unit, Public Health England — Microbiology Reference Services, Colindale, 61 Colindale Avenue, London, NW9 5EQ, United Kingdom.
**Corresponding author thomas.vaughan@phe.gov.uk*

Contents

Figure S1 Median poly(C) tract lengths of *B. pertussis* isolates UK001 to UK048, isolated from 1920 to 2012.

Isolate	ENA identifier	File URL
UK001	ERS018345	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037420/ERR037420_1.fastq.gz
UK001	ERS018345	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037420/ERR037420_2.fastq.gz
UK002	ERS018346	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037421/ERR037421_1.fastq.gz
UK002	ERS018346	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037421/ERR037421_2.fastq.gz
UK003	ERS018347	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037422/ERR037422_1.fastq.gz
UK003	ERS018347	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037422/ERR037422_2.fastq.gz
UK004	ERS018348	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037396/ERR037396_1.fastq.gz
UK004	ERS018348	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037396/ERR037396_2.fastq.gz
UK005	ERS018349	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037400/ERR037400_1.fastq.gz
UK005	ERS018349	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037400/ERR037400_2.fastq.gz
UK006	ERS018350	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037401/ERR037401_1.fastq.gz
UK006	ERS018350	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037401/ERR037401_2.fastq.gz
UK008	ERS018351	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037402/ERR037402_1.fastq.gz
UK008	ERS018351	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037402/ERR037402_2.fastq.gz
UK009	ERS018352	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037403/ERR037403_1.fastq.gz
UK009	ERS018352	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037403/ERR037403_2.fastq.gz
UK010	ERS018353	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037404/ERR037404_1.fastq.gz
UK010	ERS018353	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037404/ERR037404_2.fastq.gz
UK011	ERS018354	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037405/ERR037405_1.fastq.gz
UK011	ERS018354	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037405/ERR037405_2.fastq.gz
UK014	ERS018355	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037406/ERR037406_1.fastq.gz
UK014	ERS018355	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037406/ERR037406_2.fastq.gz
UK015	ERS018356	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037407/ERR037407_1.fastq.gz
UK015	ERS018356	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037407/ERR037407_2.fastq.gz
UK016	ERS018357	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037397/ERR037397_1.fastq.gz
UK016	ERS018357	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037397/ERR037397_2.fastq.gz
UK017	ERS018358	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037398/ERR037398_1.fastq.gz
UK017	ERS018358	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037398/ERR037398_2.fastq.gz
UK018	ERS018359	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037399/ERR037399_1.fastq.gz
UK018	ERS018359	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037399/ERR037399_2.fastq.gz

Table S1 Unassembled genome sequence files for UK isolates used in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
UK020	ERS018360	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037408/ERR037408_1.fastq.gz
UK020	ERS018360	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037408/ERR037408_2.fastq.gz
UK021	ERS018361	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037411/ERR037411_1.fastq.gz
UK021	ERS018361	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037411/ERR037411_2.fastq.gz
UK022	ERS018362	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037412/ERR037412_1.fastq.gz
UK022	ERS018362	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037412/ERR037412_2.fastq.gz
UK023	ERS018363	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037413/ERR037413_1.fastq.gz
UK023	ERS018363	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037413/ERR037413_2.fastq.gz
UK024	ERS018364	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037414/ERR037414_1.fastq.gz
UK024	ERS018364	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037414/ERR037414_2.fastq.gz
UK049	ERS227750	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304770/ERR304770_1.fastq.gz
UK049	ERS227750	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304770/ERR304770_2.fastq.gz
UK050	ERS227751	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304771/ERR304771_1.fastq.gz
UK050	ERS227751	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304771/ERR304771_2.fastq.gz
UK051	ERS227752	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304772/ERR304772_1.fastq.gz
UK051	ERS227752	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304772/ERR304772_2.fastq.gz
UK052	ERS227753	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304773/ERR304773_1.fastq.gz
UK052	ERS227753	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304773/ERR304773_2.fastq.gz
UK053	ERS227754	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304774/ERR304774_1.fastq.gz
UK053	ERS227754	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304774/ERR304774_2.fastq.gz
UK054	ERS227755	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304775/ERR304775_1.fastq.gz
UK054	ERS227755	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304775/ERR304775_2.fastq.gz
UK055	ERS227756	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304776/ERR304776_1.fastq.gz
UK055	ERS227756	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304776/ERR304776_2.fastq.gz
UK056	ERS227757	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304777/ERR304777_1.fastq.gz
UK056	ERS227757	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304777/ERR304777_2.fastq.gz
UK057	ERS227758	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304778/ERR304778_1.fastq.gz
UK057	ERS227758	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304778/ERR304778_2.fastq.gz
UK058	ERS227759	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304779/ERR304779_1.fastq.gz
UK058	ERS227759	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304779/ERR304779_2.fastq.gz

Table SI continued Unassembled genome sequence files for UK isolates used in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
UK059	ERS227760	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304780/ERR304780_1.fastq.gz
UK059	ERS227760	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304780/ERR304780_2.fastq.gz
UK060	ERS227761	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304781/ERR304781_1.fastq.gz
UK060	ERS227761	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304781/ERR304781_2.fastq.gz
UK061	ERS227762	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304782/ERR304782_1.fastq.gz
UK061	ERS227762	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304782/ERR304782_2.fastq.gz
UK062	ERS227763	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304783/ERR304783_1.fastq.gz
UK062	ERS227763	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304783/ERR304783_2.fastq.gz
UK063	ERS227764	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304784/ERR304784_1.fastq.gz
UK063	ERS227764	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304784/ERR304784_2.fastq.gz
UK064	ERS227765	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304785/ERR304785_1.fastq.gz
UK064	ERS227765	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304785/ERR304785_2.fastq.gz
UK065	ERS227766	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304786/ERR304786_1.fastq.gz
UK065	ERS227766	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304786/ERR304786_2.fastq.gz
UK066	ERS227767	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304787/ERR304787_1.fastq.gz
UK066	ERS227767	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR304/ERR304787/ERR304787_2.fastq.gz
UK084	ERS227785	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305950/ERR305950_1.fastq.gz
UK084	ERS227785	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305950/ERR305950_2.fastq.gz
UK085	ERS227786	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305951/ERR305951_1.fastq.gz
UK085	ERS227786	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305951/ERR305951_2.fastq.gz
UK086	ERS227787	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305952/ERR305952_1.fastq.gz
UK086	ERS227787	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305952/ERR305952_2.fastq.gz
UK087	ERS227788	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305953/ERR305953_1.fastq.gz
UK087	ERS227788	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305953/ERR305953_2.fastq.gz
UK088	ERS227789	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305954/ERR305954_1.fastq.gz
UK088	ERS227789	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305954/ERR305954_2.fastq.gz
UK089	ERS227790	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305955/ERR305955_1.fastq.gz
UK089	ERS227790	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305955/ERR305955_2.fastq.gz
UK090	ERS227791	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305956/ERR305956_1.fastq.gz
UK090	ERS227791	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305956/ERR305956_2.fastq.gz

Table S1 continued Unassembled genome sequence files for UK isolates used in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
UK091	ERS227792	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305957/ERR305957_1.fastq.gz
UK091	ERS227792	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305957/ERR305957_2.fastq.gz
UK092	ERS227793	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305958/ERR305958_1.fastq.gz
UK092	ERS227793	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305958/ERR305958_2.fastq.gz
UK093	ERS227794	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305959/ERR305959_1.fastq.gz
UK093	ERS227794	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305959/ERR305959_2.fastq.gz
UK094	ERS227795	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305960/ERR305960_1.fastq.gz
UK094	ERS227795	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305960/ERR305960_2.fastq.gz
UK095	ERS227796	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR30596I/ERR30596I_1.fastq.gz
UK095	ERS227796	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR30596I/ERR30596I_2.fastq.gz
UK096	ERS227797	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305962/ERR305962_1.fastq.gz
UK096	ERS227797	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305962/ERR305962_2.fastq.gz
UK097	ERS227798	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305963/ERR305963_1.fastq.gz
UK097	ERS227798	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305963/ERR305963_2.fastq.gz
UK098	ERS227799	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305964/ERR305964_1.fastq.gz
UK098	ERS227799	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305964/ERR305964_2.fastq.gz
UK099	ERS227800	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305965/ERR305965_1.fastq.gz
UK099	ERS227800	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305965/ERR305965_2.fastq.gz
UKI00	ERS22780I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305966/ERR305966_1.fastq.gz
UKI00	ERS22780I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR305/ERR305966/ERR305966_2.fastq.gz

Table SI continued Unassembled genome sequence files for UK isolates used in this study.

Isolate	ENA identifier	File URL
(n/a)	ERS001360	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007653/ERR007653_1.fastq.gz
(n/a)	ERS001360	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007653/ERR007653_2.fastq.gz
Bp02-46	ERS018336	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037419/ERR037419_1.fastq.gz
Bp02-46	ERS018336	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037419/ERR037419_2.fastq.gz
Bp03-030	ERS018337	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037423/ERR037423_1.fastq.gz
Bp03-030	ERS018337	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037423/ERR037423_2.fastq.gz
Bp03-040	ERS018338	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037424/ERR037424_1.fastq.gz
Bp03-040	ERS018338	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037424/ERR037424_2.fastq.gz
Bp06-089	ERS018339	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037425/ERR037425_1.fastq.gz
Bp06-089	ERS018339	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037425/ERR037425_2.fastq.gz
Bp06-090	ERS018340	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037426/ERR037426_1.fastq.gz
Bp06-090	ERS018340	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037426/ERR037426_2.fastq.gz
Bp06-246	ERS018341	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037427/ERR037427_1.fastq.gz
Bp06-246	ERS018341	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037427/ERR037427_2.fastq.gz
Bp06-274	ERS018342	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037428/ERR037428_1.fastq.gz
Bp06-274	ERS018342	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037428/ERR037428_2.fastq.gz
Bp06-371	ERS018343	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037429/ERR037429_1.fastq.gz
Bp06-371	ERS018343	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037429/ERR037429_2.fastq.gz
Bp06-401	ERS018344	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037430/ERR037430_1.fastq.gz
Bp06-401	ERS018344	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037430/ERR037430_2.fastq.gz
bp4376	ERS018365	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037415/ERR037415_1.fastq.gz
bp4376	ERS018365	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037415/ERR037415_2.fastq.gz
bp4389	ERS018366	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037416/ERR037416_1.fastq.gz
bp4389	ERS018366	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037416/ERR037416_2.fastq.gz
bp4397	ERS018367	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037417/ERR037417_1.fastq.gz
bp4397	ERS018367	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037417/ERR037417_2.fastq.gz
bp4396	ERS018368	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037418/ERR037418_1.fastq.gz
bp4396	ERS018368	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037418/ERR037418_2.fastq.gz
bp4398	ERS018369	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037409/ERR037409_1.fastq.gz
bp4398	ERS018369	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037409/ERR037409_2.fastq.gz

Table S2 Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
bp4307	ERS018370	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037410/ERR037410_1.fastq.gz
bp4307	ERS018370	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR037/ERR037410/ERR037410_2.fastq.gz
(n/a)	ERS045992	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044160/ERR044160_1.fastq.gz
(n/a)	ERS045992	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044160/ERR044160_2.fastq.gz
(n/a)	ERS045996	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044164/ERR044164_1.fastq.gz
(n/a)	ERS045996	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044164/ERR044164_2.fastq.gz
(n/a)	ERS046000	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044168/ERR044168_1.fastq.gz
(n/a)	ERS046000	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044168/ERR044168_2.fastq.gz
(n/a)	ERS046004	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044172/ERR044172_1.fastq.gz
(n/a)	ERS046004	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044172/ERR044172_2.fastq.gz
(n/a)	ERS046008	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044176/ERR044176_1.fastq.gz
(n/a)	ERS046008	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044176/ERR044176_2.fastq.gz
(n/a)	ERS046012	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044180/ERR044180_1.fastq.gz
(n/a)	ERS046012	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044180/ERR044180_2.fastq.gz
(n/a)	ERS046016	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044184/ERR044184_1.fastq.gz
(n/a)	ERS046016	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044184/ERR044184_2.fastq.gz
(n/a)	ERS046020	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044188/ERR044188_1.fastq.gz
(n/a)	ERS046020	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044188/ERR044188_2.fastq.gz
(n/a)	ERS046024	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044192/ERR044192_1.fastq.gz
(n/a)	ERS046024	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044192/ERR044192_2.fastq.gz
(n/a)	ERS046028	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044196/ERR044196_1.fastq.gz
(n/a)	ERS046028	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044196/ERR044196_2.fastq.gz
(n/a)	ERS046032	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044200/ERR044200_1.fastq.gz
(n/a)	ERS046032	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044200/ERR044200_2.fastq.gz
(n/a)	ERS046036	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044204/ERR044204_1.fastq.gz
(n/a)	ERS046036	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044204/ERR044204_2.fastq.gz
(n/a)	ERS046040	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044208/ERR044208_1.fastq.gz
(n/a)	ERS046040	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044208/ERR044208_2.fastq.gz
(n/a)	ERS046044	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044212/ERR044212_1.fastq.gz
(n/a)	ERS046044	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044212/ERR044212_2.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
(n/a)	ERS046048	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044216/ERR044216_1.fastq.gz
(n/a)	ERS046048	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044216/ERR044216_2.fastq.gz
(n/a)	ERS046052	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044220/ERR044220_1.fastq.gz
(n/a)	ERS046052	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044220/ERR044220_2.fastq.gz
(n/a)	ERS046056	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044224/ERR044224_1.fastq.gz
(n/a)	ERS046056	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044224/ERR044224_2.fastq.gz
(n/a)	ERS046060	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044228/ERR044228_1.fastq.gz
(n/a)	ERS046060	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044228/ERR044228_2.fastq.gz
(n/a)	ERS046064	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044232/ERR044232_1.fastq.gz
(n/a)	ERS046064	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044232/ERR044232_2.fastq.gz
(n/a)	ERS046068	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044236/ERR044236_1.fastq.gz
(n/a)	ERS046068	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044236/ERR044236_2.fastq.gz
(n/a)	ERS046072	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044240/ERR044240_1.fastq.gz
(n/a)	ERS046072	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044240/ERR044240_2.fastq.gz
(n/a)	ERS046076	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044244/ERR044244_1.fastq.gz
(n/a)	ERS046076	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044244/ERR044244_2.fastq.gz
(n/a)	ERS046080	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044248/ERR044248_1.fastq.gz
(n/a)	ERS046080	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044248/ERR044248_2.fastq.gz
(n/a)	ERS046084	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044252/ERR044252_2.fastq.gz
(n/a)	ERS046086	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044254/ERR044254_1.fastq.gz
(n/a)	ERS046086	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044254/ERR044254_2.fastq.gz
(n/a)	ERS046090	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044258/ERR044258_1.fastq.gz
(n/a)	ERS046090	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR044/ERR044258/ERR044258_2.fastq.gz
(n/a)	ERS046110	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR043/ERR043365/ERR043365_1.fastq.gz
(n/a)	ERS046110	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR043/ERR043365/ERR043365_2.fastq.gz
(n/a)	ERS046114	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR043/ERR043369/ERR043369_1.fastq.gz
(n/a)	ERS046114	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR043/ERR043369/ERR043369_2.fastq.gz
(n/a)	ERS046118	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR043/ERR043373/ERR043373_1.fastq.gz
(n/a)	ERS046118	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR043/ERR043373/ERR043373_2.fastq.gz
(n/a)	ERS046122	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR043/ERR043377/ERR043377_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
(n/a)	ERS046122	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR043/ERR043377/ERR043377_2.fastq.gz
(n/a)	ERS053508	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045441/ERR045441_1.fastq.gz
(n/a)	ERS053508	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045441/ERR045441_2.fastq.gz
(n/a)	ERS053515	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045475/ERR045475_1.fastq.gz
(n/a)	ERS053515	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045475/ERR045475_2.fastq.gz
(n/a)	ERS053517	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045438/ERR045438_1.fastq.gz
(n/a)	ERS053517	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045438/ERR045438_2.fastq.gz
(n/a)	ERS053523	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045449/ERR045449_1.fastq.gz
(n/a)	ERS053523	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045449/ERR045449_2.fastq.gz
(n/a)	ERS053525	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045453/ERR045453_1.fastq.gz
(n/a)	ERS053525	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045453/ERR045453_2.fastq.gz
(n/a)	ERS053530	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045461/ERR045461_1.fastq.gz
(n/a)	ERS053530	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045461/ERR045461_2.fastq.gz
(n/a)	ERS053535	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045471/ERR045471_1.fastq.gz
(n/a)	ERS053535	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045471/ERR045471_2.fastq.gz
(n/a)	ERS053548	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045445/ERR045445_1.fastq.gz
(n/a)	ERS053548	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045445/ERR045445_2.fastq.gz
(n/a)	ERS053552	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045457/ERR045457_1.fastq.gz
(n/a)	ERS053552	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045457/ERR045457_2.fastq.gz
(n/a)	ERS053554	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045468/ERR045468_1.fastq.gz
(n/a)	ERS053554	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR045/ERR045468/ERR045468_2.fastq.gz
(n/a)	ERS054025	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046401/ERR046401_1.fastq.gz
(n/a)	ERS054025	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046401/ERR046401_2.fastq.gz
(n/a)	ERS054026	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046406/ERR046406_1.fastq.gz
(n/a)	ERS054026	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046406/ERR046406_2.fastq.gz
(n/a)	ERS054027	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046410/ERR046410_1.fastq.gz
(n/a)	ERS054027	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046410/ERR046410_2.fastq.gz
(n/a)	ERS054028	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046414/ERR046414_1.fastq.gz
(n/a)	ERS054028	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046414/ERR046414_2.fastq.gz
(n/a)	ERS054029	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046418/ERR046418_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
(n/a)	ERS054029	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046418/ERR046418_2.fastq.gz
(n/a)	ERS054030	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046423/ERR046423_1.fastq.gz
(n/a)	ERS054030	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046423/ERR046423_2.fastq.gz
(n/a)	ERS054031	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046426/ERR046426_1.fastq.gz
(n/a)	ERS054031	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046426/ERR046426_2.fastq.gz
(n/a)	ERS054032	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046430/ERR046430_1.fastq.gz
(n/a)	ERS054032	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046430/ERR046430_2.fastq.gz
(n/a)	ERS054033	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046434/ERR046434_1.fastq.gz
(n/a)	ERS054033	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046434/ERR046434_2.fastq.gz
(n/a)	ERS054034	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046439/ERR046439_1.fastq.gz
(n/a)	ERS054034	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046439/ERR046439_2.fastq.gz
(n/a)	ERS054035	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046442/ERR046442_1.fastq.gz
(n/a)	ERS054035	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046442/ERR046442_2.fastq.gz
(n/a)	ERS054036	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046446/ERR046446_1.fastq.gz
(n/a)	ERS054036	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046446/ERR046446_2.fastq.gz
(n/a)	ERS054037	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046451/ERR046451_1.fastq.gz
(n/a)	ERS054037	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046451/ERR046451_2.fastq.gz
(n/a)	ERS054038	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046455/ERR046455_1.fastq.gz
(n/a)	ERS054038	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046455/ERR046455_2.fastq.gz
(n/a)	ERS054039	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046459/ERR046459_1.fastq.gz
(n/a)	ERS054039	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046459/ERR046459_2.fastq.gz
(n/a)	ERS054040	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046463/ERR046463_1.fastq.gz
(n/a)	ERS054040	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046463/ERR046463_2.fastq.gz
(n/a)	ERS054031	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046516/ERR046516_1.fastq.gz
(n/a)	ERS054031	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046516/ERR046516_2.fastq.gz
(n/a)	ERS005390	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019606/ERR019606_2.fastq.gz
(n/a)	ERS001339	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007510/ERR007510_1.fastq.gz
(n/a)	ERS001339	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007509/ERR007509_1.fastq.gz
(n/a)	ERS001339	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007510/ERR007510_2.fastq.gz
(n/a)	ERS001339	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007509/ERR007509_2.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
(n/a)	ERS054027	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046473/ERR046473_1.fastq.gz
(n/a)	ERS054027	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046473/ERR046473_2.fastq.gz
(n/a)	ERS001338	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007506/ERR007506_2.fastq.gz
(n/a)	ERS005388	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019603/ERR019603_2.fastq.gz
(n/a)	ERS054026	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046469/ERR046469_1.fastq.gz
(n/a)	ERS001338	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007506/ERR007506_1.fastq.gz
(n/a)	ERS054028	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046477/ERR046477_1.fastq.gz
(n/a)	ERS054029	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046481/ERR046481_1.fastq.gz
(n/a)	ERS054026	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046469/ERR046469_2.fastq.gz
(n/a)	ERS054029	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046481/ERR046481_2.fastq.gz
(n/a)	ERS054028	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046477/ERR046477_2.fastq.gz
(n/a)	ERS054033	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046524/ERR046524_1.fastq.gz
(n/a)	ERS001361	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007648/ERR007648_1.fastq.gz
(n/a)	ERS054033	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046524/ERR046524_2.fastq.gz
(n/a)	ERS001361	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR007/ERR007648/ERR007648_2.fastq.gz
(n/a)	ERS054034	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046490/ERR046490_1.fastq.gz
(n/a)	ERS054034	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046490/ERR046490_2.fastq.gz
(n/a)	ERS054039	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046510/ERR046510_1.fastq.gz
(n/a)	ERS054036	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046497/ERR046497_1.fastq.gz
(n/a)	ERS054035	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046493/ERR046493_1.fastq.gz
(n/a)	ERS054039	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046510/ERR046510_2.fastq.gz
(n/a)	ERS054036	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046497/ERR046497_2.fastq.gz
(n/a)	ERS054035	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046493/ERR046493_2.fastq.gz
(n/a)	ERS054030	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046486/ERR046486_1.fastq.gz
(n/a)	ERS054040	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046514/ERR046514_1.fastq.gz
(n/a)	ERS054030	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046486/ERR046486_2.fastq.gz
(n/a)	ERS054032	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046520/ERR046520_1.fastq.gz
(n/a)	ERS054038	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046506/ERR046506_1.fastq.gz
(n/a)	ERS054040	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046514/ERR046514_2.fastq.gz
(n/a)	ERS054032	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046520/ERR046520_2.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
(n/a)	ERS054038	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046506/ERR046506_2.fastq.gz
(n/a)	ERS054025	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046464/ERR046464_1.fastq.gz
(n/a)	ERS054025	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046464/ERR046464_2.fastq.gz
(n/a)	ERS054037	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046502/ERR046502_1.fastq.gz
(n/a)	ERS054037	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR046/ERR046502/ERR046502_2.fastq.gz
939980_2	ERS005382	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019594/ERR019594_1.fastq.gz
939980_2	ERS005382	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019594/ERR019594_2.fastq.gz
Bp02-027	ERS014547	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029992/ERR029992_1.fastq.gz
Bp02-027	ERS014547	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029992/ERR029992_2.fastq.gz
Bp02-045	ERS014550	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029984/ERR029984_1.fastq.gz
Bp02-045	ERS014550	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029984/ERR029984_2.fastq.gz
DKI2	ERS014537	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029331/ERR029331_1.fastq.gz
DKI2	ERS014537	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029331/ERR029331_2.fastq.gz
I326620_2	ERS005378	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019597/ERR019597_1.fastq.gz
I326620_2	ERS005378	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019597/ERR019597_2.fastq.gz
BP92001	ERS014612	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028888/ERR028888_1.fastq.gz
BP92001	ERS014612	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028888/ERR028888_2.fastq.gz
BP94025	ERS014615	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028891/ERR028891_1.fastq.gz
BP94025	ERS014615	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028891/ERR028891_2.fastq.gz
BII78	ERS014208	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029300/ERR029300_1.fastq.gz
BII78	ERS014208	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029300/ERR029300_2.fastq.gz
BPI82	ERS014509	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029313/ERR029313_1.fastq.gz
BPI82	ERS014509	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029313/ERR029313_2.fastq.gz
PT3783	ERS014611	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028885/ERR028885_1.fastq.gz
PT3783	ERS014611	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028885/ERR028885_2.fastq.gz
BP94035	ERS014616	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028892/ERR028892_1.fastq.gz
BP94035	ERS014616	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028892/ERR028892_2.fastq.gz
BPI28	ERS014511	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029315/ERR029315_1.fastq.gz
BPI28	ERS014511	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029315/ERR029315_2.fastq.gz
B0907	ERS014207	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029299/ERR029299_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
B0907	ERS0I4207	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029299/ERR029299_2.fastq.gz
L599	ERS0I4594	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028869/ERR028869_1.fastq.gz
L599	ERS0I4594	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028869/ERR028869_2.fastq.gz
BII4I	ERS0I42I0	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029302/ERR029302_1.fastq.gz
BII4I	ERS0I42I0	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029302/ERR029302_2.fastq.gz
(n/a)	ERS005376	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9595/ERR0I9595_1.fastq.gz
(n/a)	ERS005376	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9595/ERR0I9595_2.fastq.gz
BP70	ERS0I4507	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR0293II/ERR0293II_1.fastq.gz
BP70	ERS0I4507	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR0293II/ERR0293II_2.fastq.gz
BP56	ERS0I45I4	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029307/ERR029307_1.fastq.gz
BP56	ERS0I45I4	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029307/ERR029307_2.fastq.gz
B09I9	ERS0I42I3	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029294/ERR029294_1.fastq.gz
B09I9	ERS0I42I3	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029294/ERR029294_2.fastq.gz
BII73	ERS0I4205	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029297/ERR029297_1.fastq.gz
BII73	ERS0I4205	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029297/ERR029297_2.fastq.gz
PRCB640	ERS0I453I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029336/ERR029336_1.fastq.gz
PRCB640	ERS0I453I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029336/ERR029336_2.fastq.gz
DKI3	ERS0I4538	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029332/ERR029332_1.fastq.gz
DKI3	ERS0I4538	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029332/ERR029332_2.fastq.gz
BP3I	ERS0I45I3	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029306/ERR029306_1.fastq.gz
BP3I	ERS0I45I3	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029306/ERR029306_2.fastq.gz
L570	ERS0I4586	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028849/ERR028849_1.fastq.gz
L570	ERS0I4586	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028849/ERR028849_2.fastq.gz
B0930	ERS0I4209	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR02930I/ERR02930I_1.fastq.gz
B0930	ERS0I4209	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR02930I/ERR02930I_2.fastq.gz
PRCB52I	ERS0I4529	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029334/ERR029334_1.fastq.gz
PRCB52I	ERS0I4529	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029334/ERR029334_2.fastq.gz
BII34	ERS0I42II	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029292/ERR029292_1.fastq.gz
BII34	ERS0I42II	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029292/ERR029292_2.fastq.gz
DK7	ERS0I4533	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029338/ERR029338_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
DK7	ERS0I4533	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029338/ERR029338_2.fastq.gz
BP230	ERS0I45I8	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029322/ERR029322_1.fastq.gz
BP230	ERS0I45I8	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029322/ERR029322_2.fastq.gz
IS5840_2	ERS005370	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9587/ERR0I9587_1.fastq.gz
IS5840_2	ERS005370	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9587/ERR0I9587_2.fastq.gz
2035970_2	ERS005374	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9588/ERR0I9588_1.fastq.gz
2035970_2	ERS005374	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9588/ERR0I9588_2.fastq.gz
B09I7	ERS0I4504	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029308/ERR029308_1.fastq.gz
B09I7	ERS0I4504	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029308/ERR029308_2.fastq.gz
BP279	ERS0I45I9	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029323/ERR029323_1.fastq.gz
BP279	ERS0I45I9	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029323/ERR029323_2.fastq.gz
bp5776	ERS0I4560	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028822/ERR028822_1.fastq.gz
bp5776	ERS0I4560	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028822/ERR028822_2.fastq.gz
B088I	ERS0I862I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03II00/ERR03II00_1.fastq.gz
B088I	ERS0I862I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03II00/ERR03II00_2.fastq.gz
DKII	ERS0I4536	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029330/ERR029330_1.fastq.gz
DKII	ERS0I4536	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029330/ERR029330_2.fastq.gz
BPI20	ERS0I45I0	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029314/ERR029314_1.fastq.gz
BPI20	ERS0I45I0	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029314/ERR029314_2.fastq.gz
L590	ERS0I4592	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028867/ERR028867_1.fastq.gz
L590	ERS0I4592	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028867/ERR028867_2.fastq.gz
L582	ERS0I4590	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028865/ERR028865_1.fastq.gz
L582	ERS0I4590	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028865/ERR028865_2.fastq.gz
BPI0I	ERS0I4526	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029319/ERR029319_1.fastq.gz
BPI0I	ERS0I4526	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029319/ERR029319_2.fastq.gz
bp5766	ERS0I4557	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028830/ERR028830_1.fastq.gz
bp5766	ERS0I4557	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028830/ERR028830_2.fastq.gz
L695	ERS0I4607	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028883/ERR028883_1.fastq.gz
L695	ERS0I4607	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028883/ERR028883_2.fastq.gz
BP95003	ERS0I46I7	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028893/ERR028893_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
BP95003	ERS0146I7	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028893/ERR028893_2.fastq.gz
B0949	ERS0I8626	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I094/ERR03I094_1.fastq.gz
B0949	ERS0I8626	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I094/ERR03I094_2.fastq.gz
BPII9	ERS0145I5	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029317/ERR029317_1.fastq.gz
BPII9	ERS0145I5	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029317/ERR029317_2.fastq.gz
C900	ERS0I8609	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030013/ERR030013_1.fastq.gz
C900	ERS0I8609	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030013/ERR030013_2.fastq.gz
D492	ERS0I86I2	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030005/ERR030005_1.fastq.gz
D492	ERS0I86I2	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030005/ERR030005_2.fastq.gz
PRCB649	ERS0I4528	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029333/ERR029333_1.fastq.gz
PRCB649	ERS0I4528	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029333/ERR029333_2.fastq.gz
BI865	ERS0I8623	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03II02/ERR03II02_1.fastq.gz
BI865	ERS0I8623	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03II02/ERR03II02_2.fastq.gz
L709	ERS0I4609	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028874/ERR028874_1.fastq.gz
L709	ERS0I4609	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028874/ERR028874_2.fastq.gz
B044I	ERS0I86I6	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I095/ERR03I095_1.fastq.gz
B044I	ERS0I86I6	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I095/ERR03I095_2.fastq.gz
B0583	ERS0I86I9	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I098/ERR03I098_1.fastq.gz
B0583	ERS0I86I9	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I098/ERR03I098_2.fastq.gz
L468	ERS0I4579	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028853/ERR028853_1.fastq.gz
L468	ERS0I4579	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028853/ERR028853_2.fastq.gz
B0496	ERS0I86I8	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I097/ERR03I097_1.fastq.gz
B0496	ERS0I86I8	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I097/ERR03I097_2.fastq.gz
B0585	ERS0I8620	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I099/ERR03I099_1.fastq.gz
B0585	ERS0I8620	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I099/ERR03I099_2.fastq.gz
E495	ERS0I8625	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I093/ERR03I093_1.fastq.gz
E495	ERS0I8625	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I093/ERR03I093_2.fastq.gz
D898	ERS0I86I3	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030006/ERR030006_1.fastq.gz
D898	ERS0I86I3	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030006/ERR030006_2.fastq.gz
L580	ERS0I4588	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028863/ERR028863_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
L580	ERS0I4588	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028863/ERR028863_2.fastq.gz
B385	ERS0I8599	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030028/ERR030028_1.fastq.gz
B385	ERS0I8599	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030028/ERR030028_2.fastq.gz
FR3207	ERS0I4572	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028844/ERR028844_1.fastq.gz
FR3207	ERS0I4572	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028844/ERR028844_2.fastq.gz
L453	ERS0I4576	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028850/ERR028850_1.fastq.gz
L453	ERS0I4576	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028850/ERR028850_2.fastq.gz
FRII39	ERS0I4570	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028842/ERR028842_1.fastq.gz
FRII39	ERS0I4570	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028842/ERR028842_2.fastq.gz
BII53	ERS0I42I2	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029293/ERR029293_1.fastq.gz
BII53	ERS0I42I2	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029293/ERR029293_2.fastq.gz
L452	ERS0I4575	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028846/ERR028846_1.fastq.gz
L452	ERS0I4575	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028846/ERR028846_2.fastq.gz
FRI45	ERS0I4566	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028838/ERR028838_1.fastq.gz
FRI45	ERS0I4566	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028838/ERR028838_2.fastq.gz
L595	ERS0I4593	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028868/ERR028868_1.fastq.gz
L595	ERS0I4593	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028868/ERR028868_2.fastq.gz
BII6I	ERS0I4503	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029304/ERR029304_1.fastq.gz
BII6I	ERS0I4503	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029304/ERR029304_2.fastq.gz
L573	ERS0I4587	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028859/ERR028859_1.fastq.gz
L573	ERS0I4587	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028859/ERR028859_2.fastq.gz
D300	ERS0I86II	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR0300I5/ERR0300I5_1.fastq.gz
D300	ERS0I86II	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR0300I5/ERR0300I5_2.fastq.gz
Bp02-044	ERS0I4549	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029983/ERR029983_1.fastq.gz
Bp02-044	ERS0I4549	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029983/ERR029983_2.fastq.gz
bp5822	ERS0I4552	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028825/ERR028825_1.fastq.gz
bp5822	ERS0I4552	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028825/ERR028825_2.fastq.gz
B0935	ERS0I4202	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR02929I/ERR02929I_1.fastq.gz
B0935	ERS0I4202	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR02929I/ERR02929I_2.fastq.gz
BP9300I	ERS0I46I3	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028889/ERR028889_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
BP9300l	ERS0l46l3	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028889/ERR028889_2.fastq.gz
B0908	ERS0l4203	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029295/ERR029295_1.fastq.gz
B0908	ERS0l4203	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029295/ERR029295_2.fastq.gz
A2l4	ERS0l8624	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03l/ERR03l092/ERR03l092_1.fastq.gz
A2l4	ERS0l8624	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03l/ERR03l092/ERR03l092_2.fastq.gz
bp3092	ERS0l4553	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028826/ERR028826_1.fastq.gz
bp3092	ERS0l4553	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028826/ERR028826_2.fastq.gz
bpl773	ERS0l4556	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028829/ERR028829_1.fastq.gz
bpl773	ERS0l4556	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028829/ERR028829_2.fastq.gz
D909	ERS0l86l4	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030007/ERR030007_1.fastq.gz
D909	ERS0l86l4	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030007/ERR030007_2.fastq.gz
D075	ERS0l86l0	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR0300l4/ERR0300l4_1.fastq.gz
D075	ERS0l86l0	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR0300l4/ERR0300l4_2.fastq.gz
C670	ERS0l8607	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR0300ll/ERR0300ll_1.fastq.gz
C670	ERS0l8607	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR0300ll/ERR0300ll_2.fastq.gz
PRCB637	ERS0l4527	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029329/ERR029329_1.fastq.gz
PRCB637	ERS0l4527	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029329/ERR029329_2.fastq.gz
(n/a)	ERS005377	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR0l9/ERR0l9596/ERR0l9596_1.fastq.gz
(n/a)	ERS005377	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR0l9/ERR0l9596/ERR0l9596_2.fastq.gz
Bp02-035	ERS0l4548	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029982/ERR029982_1.fastq.gz
Bp02-035	ERS0l4548	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029982/ERR029982_2.fastq.gz
L48l	ERS0l4580	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028854/ERR028854_1.fastq.gz
L48l	ERS0l4580	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028854/ERR028854_2.fastq.gz
BP690l	ERS0l8579	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03l/ERR03l087/ERR03l087_1.fastq.gz
BP690l	ERS0l8579	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03l/ERR03l087/ERR03l087_2.fastq.gz
(n/a)	ERS005380	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR0l9/ERR0l959l/ERR0l959l_1.fastq.gz
(n/a)	ERS005380	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR0l9/ERR0l959l/ERR0l959l_2.fastq.gz
BP940l6	ERS0l46l4	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028890/ERR028890_1.fastq.gz
BP940l6	ERS0l46l4	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028890/ERR028890_2.fastq.gz
BP6	ERS0l4523	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029326/ERR029326_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
BP6	ERS0I4523	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029326/ERR029326_2.fastq.gz
B0442	ERS0I86I7	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I096/ERR03I096_I.fastq.gz
B0442	ERS0I86I7	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I096/ERR03I096_2.fastq.gz
FRI003	ERS0I4569	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR02884I/ERR02884I_I.fastq.gz
FRI003	ERS0I4569	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR02884I/ERR02884I_2.fastq.gz
2I46970_2	ERS005373	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9598/ERR0I9598_I.fastq.gz
2I46970_2	ERS005373	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9598/ERR0I9598_2.fastq.gz
606770_2	ERS005386	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9592/ERR0I9592_I.fastq.gz
606770_2	ERS005386	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9592/ERR0I9592_2.fastq.gz
Bp02-0I6	ERS0I4545	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029990/ERR029990_I.fastq.gz
Bp02-0I6	ERS0I4545	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029990/ERR029990_2.fastq.gz
L6I9	ERS0I4600	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028876/ERR028876_I.fastq.gz
L6I9	ERS0I4600	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028876/ERR028876_2.fastq.gz
Bp02-0I9	ERS0I4546	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR02999I/ERR02999I_I.fastq.gz
Bp02-0I9	ERS0I4546	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR02999I/ERR02999I_2.fastq.gz
L6I7	ERS0I4599	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028872/ERR028872_I.fastq.gz
L6I7	ERS0I4599	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028872/ERR028872_2.fastq.gz
L58I	ERS0I4589	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028864/ERR028864_I.fastq.gz
L58I	ERS0I4589	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028864/ERR028864_2.fastq.gz
bp5783	ERS0I4559	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028832/ERR028832_I.fastq.gz
bp5783	ERS0I4559	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028832/ERR028832_2.fastq.gz
DKI5	ERS0I4540	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029985/ERR029985_I.fastq.gz
DKI5	ERS0I4540	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029985/ERR029985_2.fastq.gz
C455	ERS0I8604	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030008/ERR030008_I.fastq.gz
C455	ERS0I8604	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030008/ERR030008_2.fastq.gz
B0922	ERS0I4206	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029298/ERR029298_I.fastq.gz
B0922	ERS0I4206	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029298/ERR029298_2.fastq.gz
FR743	ERS0I4568	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028840/ERR028840_I.fastq.gz
FR743	ERS0I4568	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028840/ERR028840_2.fastq.gz
C477	ERS0I8606	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR0300I0/ERR0300I0_I.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
C477	ERS0I8606	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR0300I0/ERR0300I0_2.fastq.gz
FR37I3	ERS0I4574	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028836/ERR028836_1.fastq.gz
FR37I3	ERS0I4574	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028836/ERR028836_2.fastq.gz
BPI05	ERS0I8573	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR031/ERR031081/ERR031081_1.fastq.gz
BPI05	ERS0I8573	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR031/ERR031081/ERR031081_2.fastq.gz
B490	ERS0I860I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR0300I9/ERR0300I9_1.fastq.gz
B490	ERS0I860I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR0300I9/ERR0300I9_2.fastq.gz
B3I6	ERS0I8598	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030027/ERR030027_1.fastq.gz
B3I6	ERS0I8598	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030027/ERR030027_2.fastq.gz
I758970_2	ERS00538I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR0I9/ERR0I9590/ERR0I9590_1.fastq.gz
I758970_2	ERS00538I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR0I9/ERR0I9590/ERR0I9590_2.fastq.gz
FR3407	ERS0I4573	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028835/ERR028835_1.fastq.gz
FR3407	ERS0I4573	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028835/ERR028835_2.fastq.gz
bp5640	ERS0I455I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR02882I/ERR02882I_1.fastq.gz
bp5640	ERS0I455I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR02882I/ERR02882I_2.fastq.gz
L7II	ERS0I46I0	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028875/ERR028875_1.fastq.gz
L7II	ERS0I46I0	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028875/ERR028875_2.fastq.gz
L655	ERS0I4605	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR02888I/ERR02888I_1.fastq.gz
L655	ERS0I4605	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR02888I/ERR02888I_2.fastq.gz
C467	ERS0I8605	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030009/ERR030009_1.fastq.gz
C467	ERS0I8605	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR030009/ERR030009_2.fastq.gz
BI59	ERS0I700I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR03002I/ERR03002I_1.fastq.gz
BI59	ERS0I700I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR03002I/ERR03002I_2.fastq.gz
bp6632	ERS0I4562	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028824/ERR028824_1.fastq.gz
bp6632	ERS0I4562	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028824/ERR028824_2.fastq.gz
DKI4	ERS0I4539	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR02998I/ERR02998I_1.fastq.gz
DKI4	ERS0I4539	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR02998I/ERR02998I_2.fastq.gz
C855	ERS0I8608	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR0300I2/ERR0300I2_1.fastq.gz
C855	ERS0I8608	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR030/ERR0300I2/ERR0300I2_2.fastq.gz
L5I8	ERS0I4584	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028847/ERR028847_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
L5I8	ERS0I4584	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028847/ERR028847_2.fastq.gz
Bp02-0I2	ERS0I4544	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029989/ERR029989_1.fastq.gz
Bp02-0I2	ERS0I4544	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029989/ERR029989_2.fastq.gz
BP60I	ERS0I8576	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I084/ERR03I084_1.fastq.gz
BP60I	ERS0I8576	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I084/ERR03I084_2.fastq.gz
bp890	ERS0I4554	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028827/ERR028827_1.fastq.gz
bp890	ERS0I4554	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028827/ERR028827_2.fastq.gz
BP96004	ERS0I46I9	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028895/ERR028895_1.fastq.gz
BP96004	ERS0I46I9	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028895/ERR028895_2.fastq.gz
B0403	ERS0I86I5	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I09I/ERR03I09I_1.fastq.gz
B0403	ERS0I86I5	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I09I/ERR03I09I_2.fastq.gz
B0887	ERS0I8622	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03II0I/ERR03II0I_1.fastq.gz
B0887	ERS0I8622	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03II0I/ERR03II0I_2.fastq.gz
BP95004	ERS0I46I8	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028894/ERR028894_1.fastq.gz
BP95004	ERS0I46I8	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028894/ERR028894_2.fastq.gz
BPI06	ERS0I8574	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I082/ERR03I082_1.fastq.gz
BPI06	ERS0I8574	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I082/ERR03I082_2.fastq.gz
BPI2I	ERS0I45I2	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029305/ERR029305_1.fastq.gz
BPI2I	ERS0I45I2	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029305/ERR029305_2.fastq.gz
FR3080	ERS0I457I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028843/ERR028843_1.fastq.gz
FR3080	ERS0I457I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028843/ERR028843_2.fastq.gz
PRCB450	ERS0I4530	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029335/ERR029335_1.fastq.gz
PRCB450	ERS0I4530	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029335/ERR029335_2.fastq.gz
BP3I7	ERS0I452I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029325/ERR029325_1.fastq.gz
BP3I7	ERS0I452I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029325/ERR029325_2.fastq.gz
BPI43	ERS0I45I6	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029320/ERR029320_1.fastq.gz
BPI43	ERS0I45I6	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029320/ERR029320_2.fastq.gz
bp4834	ERS0I4555	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028828/ERR028828_1.fastq.gz
bp4834	ERS0I4555	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028828/ERR028828_2.fastq.gz
B09I3	ERS0I4506	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR0293I0/ERR0293I0_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
B09I3	ERS0I4506	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR0293I0/ERR0293I0_2.fastq.gz
Bp02-00I	ERS0I454I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029986/ERR029986_I.fastq.gz
Bp02-00I	ERS0I454I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029986/ERR029986_2.fastq.gz
B225	ERS0I8597	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030026/ERR030026_I.fastq.gz
B225	ERS0I8597	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030026/ERR030026_2.fastq.gz
BP824	ERS0I8578	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I086/ERR03I086_I.fastq.gz
BP824	ERS0I8578	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I086/ERR03I086_2.fastq.gz
L508	ERS0I4583	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028857/ERR028857_I.fastq.gz
L508	ERS0I4583	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028857/ERR028857_2.fastq.gz
B2I5	ERS0I8596	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030025/ERR030025_I.fastq.gz
B2I5	ERS0I8596	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030025/ERR030025_2.fastq.gz
Bp02-005	ERS0I4543	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029988/ERR029988_I.fastq.gz
Bp02-005	ERS0I4543	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029988/ERR029988_2.fastq.gz
BII37	ERS0I4505	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029309/ERR029309_I.fastq.gz
BII37	ERS0I4505	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029309/ERR029309_2.fastq.gz
DK8	ERS0I4534	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029339/ERR029339_I.fastq.gz
DK8	ERS0I4534	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029339/ERR029339_2.fastq.gz
CI39	ERS0I8603	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030004/ERR030004_I.fastq.gz
CI39	ERS0I8603	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030004/ERR030004_2.fastq.gz
B200	ERS0I8593	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030022/ERR030022_I.fastq.gz
B200	ERS0I8593	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030022/ERR030022_2.fastq.gz
L49I	ERS0I458I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028855/ERR028855_I.fastq.gz
L49I	ERS0I458I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028855/ERR028855_2.fastq.gz
L584	ERS0I459I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028866/ERR028866_I.fastq.gz
L584	ERS0I459I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028866/ERR028866_2.fastq.gz
Bp02-002	ERS0I4542	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029987/ERR029987_I.fastq.gz
Bp02-002	ERS0I4542	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029987/ERR029987_2.fastq.gz
L535	ERS0I4585	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028848/ERR028848_I.fastq.gz
L535	ERS0I4585	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028848/ERR028848_2.fastq.gz
DK6	ERS0I4532	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029337/ERR029337_I.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
DK6	ERS0I4532	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029337/ERR029337_2.fastq.gz
I26690_2	ERS005379	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9593/ERR0I9593_1.fastq.gz
I26690_2	ERS005379	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9593/ERR0I9593_2.fastq.gz
BP706	ERS0I8577	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I085/ERR03I085_1.fastq.gz
BP706	ERS0I8577	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I085/ERR03I085_2.fastq.gz
bpI535	ERS0I456I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028823/ERR028823_1.fastq.gz
bpI535	ERS0I456I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028823/ERR028823_2.fastq.gz
L708	ERS0I4608	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028873/ERR028873_1.fastq.gz
L708	ERS0I4608	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028873/ERR028873_2.fastq.gz
BP7	ERS0I4525	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029318/ERR029318_1.fastq.gz
BP7	ERS0I4525	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029318/ERR029318_2.fastq.gz
KKKI33	ERS0I859I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030002/ERR030002_1.fastq.gz
KKKI33	ERS0I859I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030002/ERR030002_2.fastq.gz
B453	ERS0I8600	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030018/ERR030018_1.fastq.gz
B453	ERS0I8600	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030018/ERR030018_2.fastq.gz
IS8205_2	ERS005372	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9589/ERR0I9589_1.fastq.gz
IS8205_2	ERS005372	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR0I9/ERR0I9589/ERR0I9589_2.fastq.gz
B733	ERS0I8602	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030020/ERR030020_1.fastq.gz
B733	ERS0I8602	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030020/ERR030020_2.fastq.gz
Cz	ERS0I4564	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028834/ERR028834_1.fastq.gz
Cz	ERS0I4564	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028834/ERR028834_2.fastq.gz
BP22	ERS0I4508	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029312/ERR029312_1.fastq.gz
BP22	ERS0I4508	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029312/ERR029312_2.fastq.gz
BP3I2	ERS0I8575	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I083/ERR03I083_1.fastq.gz
BP3I2	ERS0I8575	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I083/ERR03I083_2.fastq.gz
L467	ERS0I4578	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028852/ERR028852_1.fastq.gz
L467	ERS0I4578	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028852/ERR028852_2.fastq.gz
PRCB406	ERS0I8590	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR03000I/ERR03000I_1.fastq.gz
PRCB406	ERS0I8590	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR03000I/ERR03000I_2.fastq.gz
A8I8	ERS0I8592	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030017/ERR030017_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
A8I8	ERS0I8592	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR0300I7/ERR0300I7_2.fastq.gz
B203	ERS0I8595	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030024/ERR030024_I.fastq.gz
B203	ERS0I8595	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030024/ERR030024_2.fastq.gz
BP2	ERS0I4524	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029327/ERR029327_I.fastq.gz
BP2	ERS0I4524	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029327/ERR029327_2.fastq.gz
Hav	ERS0I4565	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028837/ERR028837_I.fastq.gz
Hav	ERS0I4565	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028837/ERR028837_2.fastq.gz
bp6334	ERS0I4558	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR02883I/ERR02883I_I.fastq.gz
bp6334	ERS0I4558	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR02883I/ERR02883I_2.fastq.gz
B202	ERS0I8594	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030023/ERR030023_I.fastq.gz
B202	ERS0I8594	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030023/ERR030023_2.fastq.gz
DKI0	ERS0I4535	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029340/ERR029340_I.fastq.gz
DKI0	ERS0I4535	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029340/ERR029340_2.fastq.gz
BP3I0	ERS0I4520	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029324/ERR029324_I.fastq.gz
BP3I0	ERS0I4520	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029324/ERR029324_2.fastq.gz
L646	ERS0I4603	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028879/ERR028879_I.fastq.gz
L646	ERS0I4603	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028879/ERR028879_2.fastq.gz
B0934	ERS0I4204	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029296/ERR029296_I.fastq.gz
B0934	ERS0I4204	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029296/ERR029296_2.fastq.gz
PRCB352	ERS0I8589	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030000/ERR030000_I.fastq.gz
PRCB352	ERS0I8589	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR030/ERR030000/ERR030000_2.fastq.gz
L649	ERS0I4604	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028880/ERR028880_I.fastq.gz
L649	ERS0I4604	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028880/ERR028880_2.fastq.gz
L687	ERS0I4606	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028882/ERR028882_I.fastq.gz
L687	ERS0I4606	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR028882/ERR028882_2.fastq.gz
PRCBI79	ERS0I8587	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029998/ERR029998_I.fastq.gz
PRCBI79	ERS0I8587	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR029/ERR029998/ERR029998_2.fastq.gz
KKKI277	ERS0I858I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I089/ERR03I089_I.fastq.gz
KKKI277	ERS0I858I	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR03I/ERR03I089/ERR03I089_2.fastq.gz
L454	ERS0I4577	ftp://ftp.sra.ebi.ac.uk/voll/fastq/ERR028/ERR02885I/ERR02885I_I.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
L454	ERS0I4577	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR02885I/ERR02885I_2.fastq.gz
BP227	ERS0I45I7	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR02932I/ERR02932I_1.fastq.gz
BP227	ERS0I45I7	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR02932I/ERR02932I_2.fastq.gz
PRCBII8	ERS0I8586	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029997/ERR029997_1.fastq.gz
PRCBII8	ERS0I8586	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029997/ERR029997_2.fastq.gz
L6I0	ERS0I4597	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR02886I/ERR02886I_1.fastq.gz
L6I0	ERS0I4597	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR02886I/ERR02886I_2.fastq.gz
L504	ERS0I4582	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028856/ERR028856_1.fastq.gz
L504	ERS0I4582	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028856/ERR028856_2.fastq.gz
PRCB272	ERS0I8588	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029999/ERR029999_1.fastq.gz
PRCB272	ERS0I8588	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029999/ERR029999_2.fastq.gz
FR287	ERS0I4567	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028839/ERR028839_1.fastq.gz
FR287	ERS0I4567	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028839/ERR028839_2.fastq.gz
L604	ERS0I4595	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028870/ERR028870_1.fastq.gz
L604	ERS0I4595	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028870/ERR028870_2.fastq.gz
PRCB2	ERS0I8585	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029996/ERR029996_1.fastq.gz
PRCB2	ERS0I8585	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029996/ERR029996_2.fastq.gz
I9773	ERS0I8583	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029994/ERR029994_1.fastq.gz
I9773	ERS0I8583	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029994/ERR029994_2.fastq.gz
PRCB7H	ERS0I8584	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029995/ERR029995_1.fastq.gz
PRCB7H	ERS0I8584	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR029/ERR029995/ERR029995_2.fastq.gz
L638	ERS0I460I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028877/ERR028877_1.fastq.gz
L638	ERS0I460I	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028877/ERR028877_2.fastq.gz
KKK22	ERS0I8580	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I088/ERR03I088_1.fastq.gz
KKK22	ERS0I8580	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR03I/ERR03I088/ERR03I088_2.fastq.gz
L644	ERS0I4602	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028878/ERR028878_1.fastq.gz
L644	ERS0I4602	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028878/ERR028878_2.fastq.gz
L609	ERS0I4596	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028860/ERR028860_1.fastq.gz
L609	ERS0I4596	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028860/ERR028860_2.fastq.gz
(n/a)	ERS005384	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR0I9/ERR0I9600/ERR0I9600_1.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study *(continues on next page)*.

Isolate	ENA identifier	File URL
(n/a)	ERS005384	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019600/ERR019600_2.fastq.gz
L614	ERS014598	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028862/ERR028862_1.fastq.gz
L614	ERS014598	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028862/ERR028862_2.fastq.gz
3690970_2	ERS005383	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019601/ERR019601_1.fastq.gz
3690970_2	ERS005383	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019601/ERR019601_2.fastq.gz
(n/a)	ERS005387	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019604/ERR019604_1.fastq.gz
(n/a)	ERS005387	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019604/ERR019604_2.fastq.gz
(n/a)	ERS005389	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019605/ERR019605_1.fastq.gz
(n/a)	ERS005389	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019605/ERR019605_2.fastq.gz
273980_2	ERS005385	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019602/ERR019602_1.fastq.gz
273980_2	ERS005385	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019602/ERR019602_2.fastq.gz
BP96006	ERS014620	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028886/ERR028886_1.fastq.gz
BP96006	ERS014620	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028886/ERR028886_2.fastq.gz
BP9613	ERS014621	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028887/ERR028887_1.fastq.gz
BP9613	ERS014621	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR028/ERR028887/ERR028887_2.fastq.gz
PT2019_2	ERS005405	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019585/ERR019585_1.fastq.gz
PT2019_2	ERS005405	ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR019/ERR019585/ERR019585_2.fastq.gz

Table S2 (continued) Unassembled genome sequence files for non–United Kingdom isolates referred to in this study.

Isolate	Year of isolation	Serotype by agglutination	ENA identifier	Fim2				Fim3			
				No. of reads	Median poly(C) tract length	Range of poly(C) tract lengths	Median fluorescence	No. of reads	Median poly(C) tract length	Range of poly(C) tract lengths	Median fluorescence
UK001	1920	1	ERS018345	80	14	4 (12 to 15)	nd	46	9	4 (7 to 10)	nd
UK002	1942	1,2	ERS018346	30	16	10 (8 to 17)	nd	22	13	5 (12 to 16)	nd
UK003	1946	1,2,3	ERS018347	18	16	3 (15 to 17)	nd	0	nd	nd	nd
UK004	1947	1,2	ERS018348	22	15.5	2 (15 to 16)	nd	35	10	3 (9 to 11)	nd
UK005	1949	1,3	ERS018349	11	17	7 (14 to 20)	nd	13	13	9 (11 to 19)	nd
UK006	1967	1,2,3	ERS018350	5	14	2 (14 to 15)	nd	5	14	7 (13 to 19)	nd
UK008	1979	1,2	ERS018349	6	16	4 (15 to 18)	nd	19	13	9 (11 to 19)	nd
UK009	1982	1,2	ERS018350	6	15	5 (12 to 16)	nd	4	12.5	5 (9 to 13)	nd
UK010	1982	1,2,3	ERS018353	1	17	1 (17 to 17)	nd	1	15	1 (15 to 15)	nd
UK011	1983	1,3	ERS018354	12	14	8 (11 to 18)	nd	2	16.5	8 (13 to 20)	nd
UK014	1998	1,2	ERS018355	16	17	10 (8 to 17)	nd	23	9	1 (9 to 9)	nd
UK015	2002	1,3	ERS018356	38	9	1 (9 to 9)	nd	1	14	1 (14 to 14)	nd
UK016	2006	1,2	ERS018357	20	17	3 (15 to 17)	nd	27	9	4 (7 to 10)	nd
UK017	2007	1,3	ERS018358	66	10	6 (6 to 11)	nd	1	14	1 (14 to 14)	nd
UK018	2010	1,3	ERS018359	37	10	2 (9 to 10)	nd	7	19	7 (14 to 20)	nd
UK020	2009	1,3	ERS018360	83	9	2 (9 to 10)	nd	5	15	4 (13 to 16)	nd
UK021	2010	1,2	ERS018361	5	14	4 (13 to 16)	nd	34	12	3 (11 to 13)	nd
UK022	2010	1,2	ERS018362	7	15	5 (12 to 16)	nd	35	11	3 (10 to 12)	nd
UK023	2010	1,3	ERS018363	92	10	3 (9 to 11)	nd	11	15	9 (12 to 20)	nd
UK024	2010	1,3	ERS018364	99	10	3 (9 to 11)	nd	11	15	8 (13 to 20)	nd
UK049	2012	1,3	ERS227750	51	10	3 (9 to 11)	nd	2	15	3 (14 to 16)	nd
UK050	2012	1,2	ERS227751	15	16	6 (13 to 18)	nd	21	10	3 (9 to 11)	nd
UK051	2011	1,3	ERS227752	48	12	3 (11 to 13)	nd	6	13	3 (12 to 14)	nd
UK052	2011	1,2	ERS227753	6	13.5	6 (10 to 15)	nd	22	11	3 (10 to 12)	nd
UK053	2011	1,3	ERS227754	37	10	3 (9 to 11)	nd	5	14	3 (13 to 15)	nd
UK054	2011	1,3	ERS227755	18	10	2 (10 to 11)	nd	3	14	2 (13 to 14)	nd
UK055	2012	1,2	ERS227756	8	16.5	7 (12 to 18)	nd	31	10	3 (9 to 11)	nd
UK056	2012	1,2	ERS227757	15	14	4 (12 to 15)	nd	15	10	2 (9 to 10)	nd
UK057	2012	1,3	ERS227758	48	9	2 (8 to 9)	nd	5	15	13 (7 to 19)	nd
UK058	2012	1,2	ERS227759	12	15	7 (12 to 18)	nd	14	12	3 (11 to 13)	nd
UK059	2012	1,3	ERS227760	63	10	3 (9 to 11)	nd	12	14.5	8 (12 to 19)	nd
UK060	2012	1,2	ERS227761	9	16	4 (13 to 16)	nd	18	12	4 (10 to 13)	nd
UK061	2012	1,3	ERS227762	36	11	3 (10 to 12)	nd	2	11	9 (7 to 15)	nd
UK062	2012	1,2	ERS227763	9	15	5 (13 to 17)	nd	19	11	4 (10 to 13)	nd
UK063	2012	1,3	ERS227764	45	10	2 (9 to 10)	nd	5	13	5 (11 to 15)	nd
UK064	2012	1,3	ERS227765	27	14	8 (9 to 16)	nd	9	15	7 (13 to 19)	nd
UK065	2012	1,3	ERS227766	62	10	4 (8 to 11)	nd	12	14	8 (13 to 20)	nd
UK066	2012	1,3	ERS227767	50	9	2 (9 to 10)	nd	14	13	6 (11 to 16)	nd
UK067	2012	1,2	ERS227768	9	15	3 (14 to 16)	nd	33	12	4 (10 to 13)	nd
UK068	2012	1,2,3	ERS227769	13	15	8 (9 to 16)	nd	11	14	7 (14 to 20)	nd
UK069	2012	1,3	ERS227770	41	11	4 (9 to 12)	nd	11	13	2 (13 to 14)	nd
UK070	2012	1,3	ERS227771	42	11	4 (10 to 13)	nd	16	13	7 (9 to 15)	nd
UK071	2012	1,2	ERS227772	7	14	8 (9 to 16)	nd	12	13	3 (12 to 14)	nd
UK072	2012	1,3	ERS227773	39	12	4 (10 to 13)	nd	14	13.5	3 (13 to 15)	nd
UK073	2012	1,2	ERS227774	14	15	7 (12 to 18)	nd	19	12	3 (11 to 13)	nd
UK074	2012	1,3	ERS227775	45	9	3 (8 to 10)	nd	11	14	9 (11 to 19)	nd
UK075	2012	1,3	ERS227776	44	10	2 (9 to 10)	nd	11	14	7 (11 to 17)	nd
UK076	2012	1,2	ERS227777	10	15	6 (12 to 17)	nd	53	12	12 (7 to 18)	nd
UK077	2012	1,3	ERS227778	53	10	3 (9 to 11)	nd	15	14	3 (13 to 15)	nd
UK078	2012	1,2	ERS227779	12	15.5	5 (12 to 16)	nd	38	12	11 (10 to 20)	nd
UK079	2012	1,2	ERS227780	14	14.5	5 (12 to 16)	nd	29	12	4 (11 to 14)	nd
UK080	2012	1,3	ERS227781	33	9	1 (9 to 9)	nd	8	14.5	10 (6 to 15)	nd
UK081	2012	1,3	ERS227782	13	16	5 (13 to 17)	nd	81	9	3 (8 to 10)	nd
UK082	2012	1,3	ERS227783	41	10	3 (9 to 11)	nd	11	14	10 (6 to 15)	nd
UK083	2012	1,2	ERS227784	9	16	3 (15 to 17)	nd	51	11	4 (9 to 12)	nd
UK084	2012	1,2	ERS227785	7	15	4 (13 to 16)	nd	34	10	3 (9 to 11)	nd
UK085	2012	1,3	ERS227786	44	10	5 (7 to 11)	nd	9	14	8 (13 to 20)	nd
UK086	2012	1,2	ERS227787	9	16	11 (7 to 17)	nd	26	10	3 (9 to 11)	nd
UK087	2012	1,3	ERS227788	37	9	3 (8 to 10)	nd	12	14	4 (12 to 15)	nd
UK088	2012	1,3	ERS227789	29	9	3 (8 to 10)	nd	5	14	3 (13 to 15)	nd
UK089	2012	1,3	ERS227790	35	10	4 (9 to 12)	nd	16	14	4 (12 to 15)	nd
UK090	2012	1,3	ERS227791	27	10	3 (9 to 11)	nd	12	14	5 (13 to 17)	nd
UK091	2012	1,3	ERS227792	29	11	3 (10 to 12)	nd	5	15	4 (13 to 16)	nd
UK092	2012	1,2	ERS227793	11	15	11 (6 to 16)	nd	41	11	4 (10 to 13)	nd
UK093	2012	1,3	ERS227794	33	10	3 (9 to 11)	nd	10	15	10 (7 to 16)	nd
UK094	2012	1,2	ERS227795	4	16	3 (14 to 16)	nd	47	10	3 (9 to 11)	nd
UK095	2012	1,3	ERS227796	47	10	3 (9 to 11)	nd	16	14	13 (6 to 18)	nd
UK096	2012	1,2	ERS227797	4	7.5	3 (6 to 8)	nd	0	nd	nd	nd
UK097	2012	1,3	ERS227798	41	10	3 (9 to 11)	nd	21	14	7 (10 to 16)	nd
UK098	2012	1,3	ERS227799	24	10	3 (9 to 11)	nd	14	14	13 (6 to 18)	nd
UK099	2012	1,2	ERS227800	9	14	11 (6 to 16)	nd	27	11	5 (9 to 13)	nd
UK100	2012	1,3	ERS227801	38	10	2 (10 to 11)	nd	10	14	4 (12 to 15)	nd

Table S3 Promoter poly(C) tract lengths of *fim*2 & *fim*3, and serotypes by agglutination, of *B. pertussis* isolates UK001 to UK024 and UK049 to UK100. ENA identifiers are European Nucleotide Archive sample names. Serotype by agglutination uses the nomenclature of Robinson *et al.* (1989b). Surface fluorescence was not determined (nd). Values (including fluorescence) for the other 22 UK isolates are in Table 2.